

MODELO PREDICTIVO PARA INFERIR EN EL PRÓXIMO PRESIDENTE DE
ESTADO A TRAVÉS DE UN VOCABULARIO ONTOLOGICO EN TWITTER

IVAN MATEO CASTAÑO GOMEZ

UNIVERSIDAD CATÓLICA DE COLOMBIA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTA D.C
2019

MODELO PREDICTIVO PARA INFERIR EN EL PRÓXIMO PRESIDENTE DE
ESTADO A TRAVÉS DE UN VOCABULARIO ONTOLOGICO EN TWITTER

IVAN MATEO CASTAÑO GOMEZ

TRABAJO INVESTIGATIVO PARA ADQUIRIR EL GRADO DE INGENIERO DE
SISTEMAS

JOHN ALEXANDER VELANDIA VEGA
INGENIERO DE SISTEMAS DE LA UNIVERSIDAD

UNIVERSIDAD CATÓLICA DE COLOMBIA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTA D.C
2019



Atribución-NoComercial 2.5 Colombia (CC BY-NC 2.5 CO)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Advertencia](#)

Usted es libre para:



Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y crear a partir del material

El licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe darle crédito a esta obra de manera adecuada, proporcionando un enlace a la licencia, e indicando si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo del licenciante.



NoComercial — Usted no puede hacer uso del material con finés comerciales.

No hay restricciones adicionales — Usted no puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otros hacer cualquier uso permitido por la licencia.

Aviso:

Usted no tiene que cumplir con la licencia para los materiales en el dominio público o cuando su uso esté permitido por una excepción o limitación aplicable.

No se entregan garantías. La licencia podría no entregarle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como relativos a publicidad, privacidad, o derechos morales pueden limitar la forma en que utilice el material.

Aprobado por el comité de grado
en cumplimiento con los
requisitos exigidos por la Facultad
de Ingeniería de la Universidad
Católica de Colombia para optar
el título de Ingeniero de Sistemas.

Ing. John Velandia
Director

Firma del jurado

Bogotá, 29 de abril del 2019

A mis padres y familiares, por su apoyo incondicional, a mis abuelos maternos porque siempre anhelaron verme profesional y hoy con orgullo puedo decir que alcance una doble ingeniería.

CONTENIDO

1. INTRODUCCIÓN	13
1.1 ANTECEDENTES	14
1.2 PLANTEAMIENTO DEL PROBLEMA	16
1.2.1 DESCRIPCIÓN DEL PROBLEMA	16
1.2.2 FORMULACIÓN DEL PROBLEMA	17
1.3 OBJETIVOS	18
1.3.1 OBJETIVO GENERAL	18
1.3.2 OBJETIVOS ESPECÍFICOS	18
1.4 JUSTIFICACIÓN	19
1.5 DELIMITACIÓN	20
1.5.1 ESPACIO	20
1.5.2 TIEMPO	20
1.5.3 CONTENIDO	20
1.5.4 ALCANCE	20
1.6 MARCO DE REFERENCIA	21
1.6.1 MARCO TEÓRICO	21
1.6.2 MARCO CONCEPTUAL.	22
2. DISEÑO METODOLÓGICO	23
3. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 1 – DEFINIR EL MÉTODO PARA LA EXTRACCIÓN DE DATOS ENLAZADOS PARA ENCONTRAR LA INFORMACIÓN	25
3.1 ANÁLISIS Y DISEÑO CONCEPTUAL DE LA EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS	25
3.2 MODELO ONTOLOGICO QUE REPRESENTA TENDENCIAS POLITICAS	27
4. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 2 – ESTABLECER UN MODELO PREDICTIVO QUE EVIDENCIE UN RESULTADO CERCANO A LA REALIDAD DE LOS DATOS EXTRAIDOS	36

4.1 MODELOS PREDICTIVOS QUE APLICAN PARA EL ANALISIS MEDIANTE UN VOCABULARIO ONTOLOGICO	36
4.2 MODELO PREDICTIVO PARA ONTOLOGIAS MEDIANTE EL ALGORITMO DE ARBOL DE DECISIONES	38
5. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 3 – DISEÑAR UNA ARQUITECTURA DE SOFTWARE QUE CUMPLA CON REQUISITOS FUNCIONALES Y NO FUNCIONALES PARA LA IMPLEMENTACIÓN DEL MODELO.	40
5.1 ARQUITECTURA DE SOFTWARE	40
6. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 4 – DESARROLLAR UNA PRUEBA DE CONCEPTO, BASADA EN LA ARQUITECTURA ESTABLECIDA.	43
6.1 PROTOTIPO DE APLICACIÓN MOVIL.	43
7. ANÁLISIS DE RESULTADOS	50
8. CONCLUSIONES	53
9. TRABAJOS FUTUROS	54
BIBLIOGRAFIA	55
ANEXOS	57

LISTA DE TABLAS

Tabla 1. Vocabulario ontológico	41
Tabla 2. Técnicas de regresión	45
Tabla 3. Técnicas de aprendizaje automático	46
Tabla 4. Resultados obtenidos mediante el árbol de decisión	61
Tabla 5. Resultados obtenidos mediante el árbol de decisión	61
Tabla 6. Matriz de confusión	62

LISTA DE ILUSTRACIONES

Ilustración 1. Evolución de la web	21
Ilustración 2. Capas de la web semántica	22
Ilustración 3. Proceso de minería de datos	23
Ilustración 4. Diagrama árbol de decisión	24
Ilustración 5. Algoritmo ID3.....	25
Ilustración 6. Uso de redes sociales en Colombia.....	27
Ilustración 7. Uso de Redes Sociales en Colombia.....	28
Ilustración 8. Conexión entre cliente y api twitter	35
Ilustración 9. Script de filtro de información.....	36
Ilustración 10. Ontología propuesta	40
Ilustración 11. Script de análisis mediante el árbol de decisión	48
Ilustración 12. Modelo predictivo.....	49
Ilustración 13. Diagrama general de los componentes involucrados para la solución propuesta.....	50
Ilustración 14. Vista de despliegue de la aplicación móvil	52
Ilustración 15. Pantalla de autenticación	54
Ilustración 16. Pantalla de registro	55
Ilustración 17. Dashboard	56
Ilustración 18. Tweets en tiempo real.....	57
Ilustración 19. Información del autor	58
Ilustración 20. Alertas	59

LISTA DE ECUACIONES

Ecuación 1. Formula de muestro	60
Ecuación 2. Exactitud del algoritmo	62

LISTA DE ANEXOS

Anexo A. Componentes del BackEnd	67
Anexo B. Datos de entrenamiento mediante ontologias	68
Anexo C. Componentes del frontend	69
Anexo D. Vista de ionic framework	70

GLOSARIO

TWITTER: Red social que permite relacionarse con otros usuarios mediante mensajes en la misma red.

TWEET: Conjunto de caracteres que no sobrepasa los 280, los cuales expresan las opiniones de los usuarios de twitter.

ETL: Con sus siglas Extracción – Transformación y Carga, es un método de extracción de datos, el cual permite manipular los datos que se encuentran en formatos diferentes para ser cargados.

POC: Con sus siglas en ingles Proof of Concept – es una implementación, que permite medir resultados en tiempos cortos para de esta forma desarrollarla completa.

ONTOLOGÍA: Es un vocabulario de un tema determinado en específico, el cual conlleva términos básicos y relaciones entre los mismos.

APLICACIÓN MOVÍL HÍBRIDA: Estas son desarrolladas con las bases de aplicaciones web, pero manejan un framework, el cual permite adaptar la vista a cualquier dispositivo y sistemas operativos móviles.

STREAMING: Con su significado en español retransmisión, en donde el usuario se enfoca en usar los datos o el producto el cual está conectado a la red y a su vez lo va descargando.

REPLY: Permite a los usuarios de twitter escribir respuestas a un tweet que ya ha sido publicado por otro usuario o el mismo usuario puede dejar un mensaje a su tweet anteriormente publicado.

TOKEN: Permite facilitar la autenticación de los usuarios.

RESUMEN

La constante búsqueda de tendencias políticas y favorabilidad, ha convertido la web en la plataforma social e interactiva más visitada. Permitiendo a las personas comunicarse entre ellos y aportar ideas a estas plataformas, generando cantidad de datos que pueden ser empleados para distintos usos como lo son la inteligencia de negocios.

Mucha de esta información no es relevante o simplemente no es importante para el usuario que la genera, pero existen formas de sacarle provecho a esta información ya sea política, financiera, educativa entre otros, permitiendo conectar cada tema de acuerdo a su contenido, evitando duplicidad de información y reutilización la misma.

El presente artículo está enfocado en el ámbito político, referente a la publicaciones realizadas por los usuarios respecto a los personajes más influyentes en Colombia en la política, clasificando estos mensajes en tres categorías: “Positivo”, “Neutro”, “Negativo”, generando un favoritismo político y una tendencia política para cada personaje seleccionado, con el objetivo de tener una visión clara respecto al favoritismo en una posibles elecciones presidenciales, implementando unos principios básicos de vocabularios ontológicos y un modelo predictivo que permitirá analizar cada mensaje que es publicado en tiempo real y clasificarlo según la categoría anteriormente mencionada.

De acuerdo a lo expuesto anteriormente, se plantea brindar una alternativa como lo es una aplicación móvil híbrida, la cual tiene la finalidad de que cualquier persona pueda ver en tiempo real las opiniones que se generan en la Twitter de su personaje favorito, desde la comodidad de su dispositivo móvil, en cualquier lugar teniendo una conexión a internet.

PALABRAS CLAVE: MODELO PREDICTIVO, APLICACIÓN MOVIL, ONTOLOGIAS, ANÁLISIS DE SENTIMIENTO, ARBOLES DE DECISIÓN, TWITTER, TWEET.

1. INTRODUCCIÓN

Durante los últimos años en Colombia hemos podido observar, como las firmas encuestadoras han incluido en internet y los medios sociales, ya que muchas implementan encuestas vías telefónicas y han ido perdiendo factibilidad y eficiencia media estas, por lo tanto, recurren a generar encuestas mediante plataformas digitales, más específicamente en redes sociales, pero muchas de estas encuestas han sido obsoletas o en el peor de los casos no representan un acercamiento a los resultados, según el periódico Latin America post se han perdido bastante credibilidad en las encuestas acerca del intención del voto, ya que muchas de estas no son respondidas a consciencia o varios de los datos son inventados por los encuestadores, en el peor de los casos la implementación de robots que responden estas encuestas y no cumplen con la función asertiva¹. Este documento, emplea tecnologías innovadoras como lo son algoritmos de predicción y un adecuado vocabulario ontológico, el cual permitirá entrenar estos algoritmos y podrá procesar mensajes de Twitter, con el fin de catalogar cada tweet según el sentimiento de la persona y clasificarlo en un candidato político respectivo.

Se pretende usar los datos generados por los usuarios miembros de la red social Twitter, los cuales no se tienen en cuenta por las firmas encuestadoras las cuales son contratadas por entes políticos para predecir con anterioridad que candidato alcanzara la presidencia, así mismo los resultados obtenidos no tienen margen de manipulación, puesto que el modelo diseñado, tratara los datos para obtener valores que arrojaran una predicción, la cual podrán inferir de manera contundente quien podría ser el próximo presidente electo en el nicho población establecido.

¹ GUERRERA EDWIN. ¿Por qué las encuestas tradicionales de intención de voto no son tan creíbles? [en línea]. < <https://latinamericanpost.com/es/20894-colombia-por-que-las-encuestas-tradicionales-de-intencion-de-voto-no-son-tan-creibles> > [citado el 26 de abril del 2019]

1.1 ANTECEDENTES

1.2 Políticos.

El caso de estudio “Candidate Reputation Management PR Campaign”, creado por “Cambridge Analytic”, tenía como eje principal apoyar la candidatura del alcalde de Bogotá el Dr. Enrique Peñalosa durante su periodo de campaña electoral 2015, la compañía realizó una investigación la cual dio como resultado que los posibles votantes presentaban desconfianza y que era improbable ganar las elecciones, por tanto como recomendación indicaron que debían presentar personas que a percepción de los votantes fueran limpias y así el mensaje que estas emitieran sería bien recibido. (Cambridge Analytica, 2018), finalmente, y gracias a este estudio el Dr. Enrique Peñalosa ganó las votaciones del año 2015 y fue nombrado alcalde de Bogotá para el periodo 2016-2019.

¿Quién será el Presidente de Colombia en 2018?, creado por la compañía “Resolve Studio” en el año 2017 y lanzado al público en el año 2018; busca predecir comportamientos e intenciones de voto con aplicación a marcas, categorías y política en una situación especial como lo es el proceso electoral colombiano. Esto se logra *“Tomando más de 12 millones de datos recogidos de redes sociales, patrones de búsqueda y otras fuentes para construir el Índice de Intención de Voto”* ... (S.A.S., RESOLVE STUDIO, 2017) Utilizando inteligencia artificial, desarrollaron un algoritmo basado en información, acciones de RRSS y en tendencias de búsquedas. en donde se toman los datos semanalmente y se calcula las predicciones con base en todas las reacciones espontáneas de las personas sin el sesgo de un encuestador o de una muestra demográfica.

El caso de estudio “Donald J. Trump for President” que fue creado por la compañía “Cambridge Analytic” alrededor de 2016. Al analizar millones de puntos de datos, identificaron consistentemente a los votantes más persuasivos y los problemas que más les interesaban, terminaron enviando mensajes específicos en horarios claves a redes sociales, YouTube y buscadores digitales, de esta forma dando a conocer las propuestas del candidato. Este marketing digital generó que el apoyo en ese momento al candidato a la presidencia de los Estados Unidos Donald J. Trump creciera y finalmente incentivara a los votantes claves a salir el día de las elecciones. Mediante la Ciencia de Datos, construyeron 20 modelos de datos personalizados para predecir el comportamiento de los votantes. Esto finalmente contribuyó a la extraordinaria victoria de Donald Trump en las elecciones presidenciales de 2016. (Cambridge Analytica, 2018).

A continuación, se citan opiniones de los medios de comunicación acerca de la influencia de las redes sociales en la política:

- *“en los últimos 7 años se ha generado un afianzamiento de la big data, que es la posibilidad de poder acceder a datos reales de votantes reales que va más allá del impacto métrico de las redes sociales”, “lo que permite que muchos de estos posibles votantes se puedan contactar a través de diferentes sistemas y conocer una respuesta directa y una medición de su accionar”. (ELTIEMPO.COM, 2018).*
- *“Con este panorama las campañas políticas que podrían transformar más ‘likes’ en votos serán, según los expertos, las que se esfuercen por realizar una buena segmentación de públicos y construir discursos que se dirijan a cada segmento poblacional, en donde se tenga en cuenta los intereses de los jóvenes y sus aportes para la construcción de programas de gobierno”. (ELTIEMPO.COM, 2018).*
- *“de acuerdo con un estudio publicado este año por Co Internet y el Centro Nacional de Consultoría, las redes sociales son el segundo medio que eligen los votantes para informarse en época electoral, detrás de los noticieros de televisión. El 57 % de los interrogados indicó que acude a las redes para enterarse de las noticias relacionadas con las elecciones y las propuestas de los candidatos. (EL COLOMBIANO, 2018).*

1.2 PLANTEAMIENTO DEL PROBLEMA

1.2.1 Descripción del problema.

La denominada “Ola Verde”, fenómeno de creatividad y apoyo que se generó en favor del entonces candidato a la presidencia el Dr. Antanas Mockus propagado por Facebook y, en menor medida, en Twitter, en el año 2010. En el cual, este movimiento generado en redes sociales no alcanzo sus objetivos, debido a que, Juan Manuel Santos impuso su victoria con 9,028,943 superando los 3,587,975 de votos en segunda vuelta del Dr. Mockus. Sin embargo, desde el año 2010 al presente el uso de redes se ha masificado e intensificado, así mismo los candidatos a los diferentes puestos políticos son activos en las RRSS, lo cual ayuda a formar un nicho poblacional que fortalece su posicionamiento e imagen frente a los usuarios de estas plataformas digitales que finalmente son ciudadanos claves “con posibilidad de sufragar”. (El Espectador & Universidad Sergio Arboleda, 2018).

La opinión de los consumidores de redes sociales, de los políticos que aspiran a cargos de altos mandos y de quienes están en el poder es fundamental puesto que cada uno de los autores genera temas y contenido que a su vez ocasiona percepciones que se asocian directamente a reacciones y que finalmente evidencian intenciones de voto de un nicho poblacional que se encuentra en redes sociales, sin embargo la “Ola Verde” para el caso de Colombia demarca un precedente puesto que indica que los comentarios, reacciones entre otros derivados de RRSS no necesariamente se verá reflejado en votos.

Un ejemplo del uso de redes sociales en política satisfactorio y el impacto que puede tener sobre los internautas, se generó con las elecciones presidenciales de Estados Unidos para el año 2016, en este caso quien ahora es el actual presidente Donald Trump superaba en seguidores a su contrincante Hilary Clinton, el Republicano contaba con 9.4% en Facebook y 7.5% en Twitter de seguidores versus la Demócrata quien tenía 4.5% y 7.5% respectivamente. Como se puede investigar hoy en día el Presidente Trump uso sus redes sociales para mover masas, puesto que en su mayoría publicaba tuits los cuales podían generar tendencia mundial en RRSS que finalmente eran estrategias para impulsar su campaña hacia la presidencia. (CNN Español, 2016).

El “MINTIC” realizo un estudio para determinar que tanto usan los colombianos internet en el año 2017. El estudio se realizó en 96 municipios a 8.300 ciudadanos entre 16 y 70 años, se concluye que el 80% de los encuestados usa Facebook y que el 20% usa Twitter. (CARACOL RADIO, 2017). Con esto el nicho poblacional podrá ser establecido y ya que se aproximan elecciones presidenciales todos aquellos que apuntan por el poder aumentaron su crecimiento en RRSS, con los cual se generan menciones positivas y negativas.

Firmas encuestadoras tales como (Invamer, 2016), o (GUARUMO, 2018) no contemplan metodologías de investigación en las cuales se utilice los comentarios y reacciones generadas por usuarios en redes sociales para inferir un posible ganador en elecciones políticas.

Actualmente firmas encuestadoras utilizan como principal técnica de recolección las encuestas presenciales, este método puede ocasionar manipulación de la información. Esta investigación emplea un vocabulario ontológico en donde se realiza: extracción, transformación, carga e interpretación de la información con índole político publicada en la red social twitter con el fin de presentar resultados cuyo origen no puede ser manipulado y así inferir el próximo presidente de estado en Colombia.

1.2.2 Formulación del problema.

De acuerdo al problema expuesto se plantea la siguiente pregunta de investigación: ¿Cómo inferir el próximo presidente en Colombia utilizando un modelo predictivo basado en un vocabulario ontológico?

1.3 OBJETIVOS

1.3.1 Objetivo general.

- Determinar un modelo que permita predecir el próximo presidente de estado colombiano, basado en el contenido generado en Twitter utilizando un vocabulario ontológico.

1.3.2 Objetivos específicos.

1. Definir el método para la extracción de datos enlazados para encontrar la información relevante a procesar.
2. Establecer un modelo predictivo que permita determinar el próximo presidente.
3. Diseñar una arquitectura de software que cumpla con requisitos funcionales y no funcionales para la implementación del modelo.
4. Desarrollar una prueba de concepto, basado en la arquitectura establecida.

1.4 JUSTIFICACIÓN

En principio, todo el contenido generado por internautas en redes sociales, se encuentra disperso y variado, en el cual, se evidencian distintas posturas en cuanto a su preferencia electoral, tendencias políticas y opiniones. De manera que, la información hacia una posición electoral es dispersa y no conlleva a un posible resultado electoral, permitiendo no tener un claro panorama electoral para las elecciones presidenciales. Por lo tanto, la implementación de un vocabulario ontológico y un procesamiento adecuado de este, abre el camino para poder relacionar datos y que estos den la posible solución a lo planteado a un acercamiento del sentimiento del usuario hacia su personaje político.

No obstante, el contenido político que se genera en redes sociales, puede alimentar distintos campos políticos que no son correctamente analizados y conllevados a un posible resultado, dando de este modo a dar datos erróneos y no lo suficientemente verídicos como para poder dar una respuesta al candidato electo. Por lo tanto, mediante la implementación de un algoritmo y un adecuado vocabulario ontológico el cual es el insumo para el entrenamiento del algoritmo permitirá exponer la información, analizarla y conlleva a una respuesta orientada a las tendencias políticas en redes sociales, en la cual podrá ser visualizada mediante una experiencia de usuario móvil híbrida, esta pueda ser accedida desde cualquier dispositivo de una forma limpia y clara.

1.5 DELIMITACIÓN

1.5.1 Espacio.

Este proyecto se desarrolló en las instalaciones de la Universidad Católica de Colombia, en la búsqueda de un modelo predictivo que permita llegar a un resultado casi idéntico al de las elecciones presidenciales en Colombia del año 2018.

1.5.2 Tiempo.

En principio, este proyecto se inició con las indagaciones electorales presidenciales en Colombia del año 2018, en el cual se llevó al cuestionamiento de la implementación de un modelo predictivo asertivo para dichas elecciones, acto seguido, en el año 2019 del primer semestre se inicia el camino hacia el desarrollo investigativo que plante dicha indagación, con un tiempo aproximado de 4 meses para la sustentación del mismo.

1.5.3 Contenido.

Inicialmente el presente documento contiene la investigación exhaustiva de dicho modelo y evalúa el comportamiento de este algoritmo mediante streaming en Twitter evaluando en tiempo real los tweets publicados por los usuarios de la red social.

1.5.4 Alcance.

Realizar una prueba de concepto móvil, que permita evidenciar el streaming de Twitter y el funcionamiento del algoritmo seleccionado.

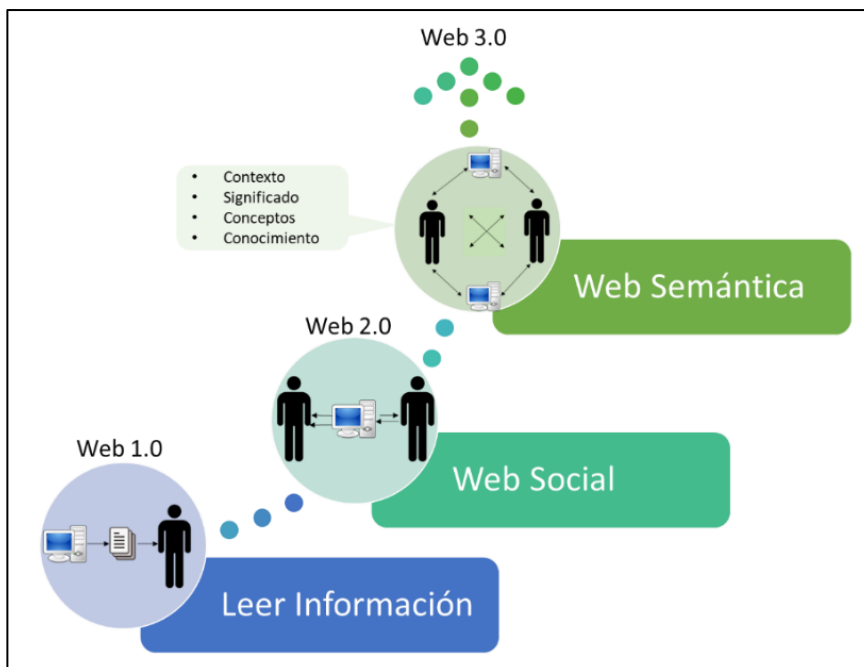
1.6 MARCO DE REFERENCIA

1.6.1 Marco Teórico.

1.6.1.1 Web Semántica.

La web semántica es la extensión de la web actual, con la cual la información obtiene una definición y un significado, conectado de esta forma los computadores y las personas que trabajan en este, cooperando e interrelacionándose unos con otros. (BERNERS-LEE, 2001). Mediante la web semántica se pueden crear grandes almacenes de datos, vocabularios y/o redactar reglas para manejar los mismos, los cuales deben estar potenciados por tecnologías como RDF, SPARQL, OWL y SKOS. (W3C, 2015).

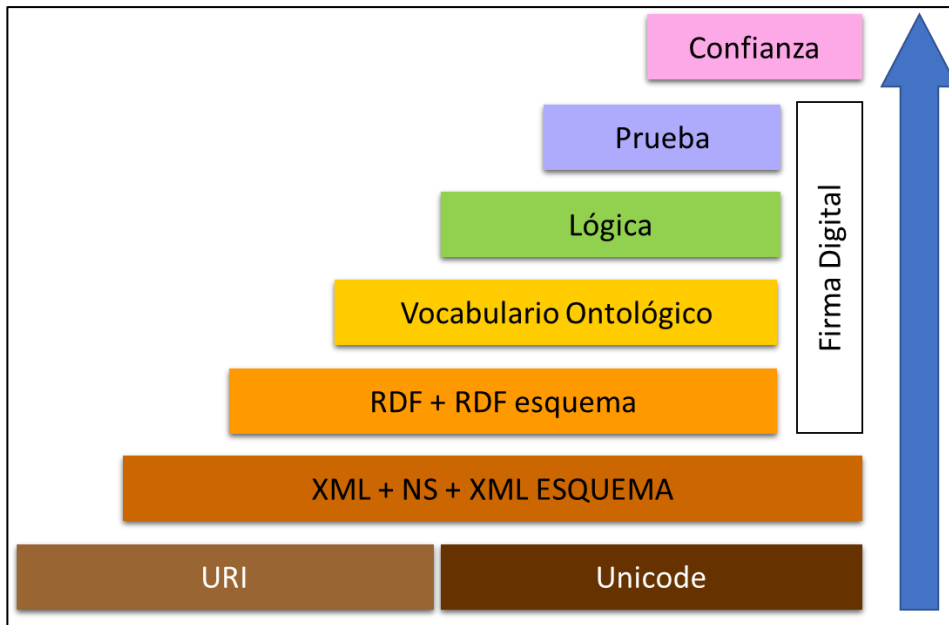
Ilustración 1. Evolución de la web



FUENTE: (Guerrero, 2018)

Según (BERNERS-LEE, 2001) para poder implementar la web semántica se puede generar una esquematización según la imagen a continuación:

Ilustración 2. Capas de la web semántica

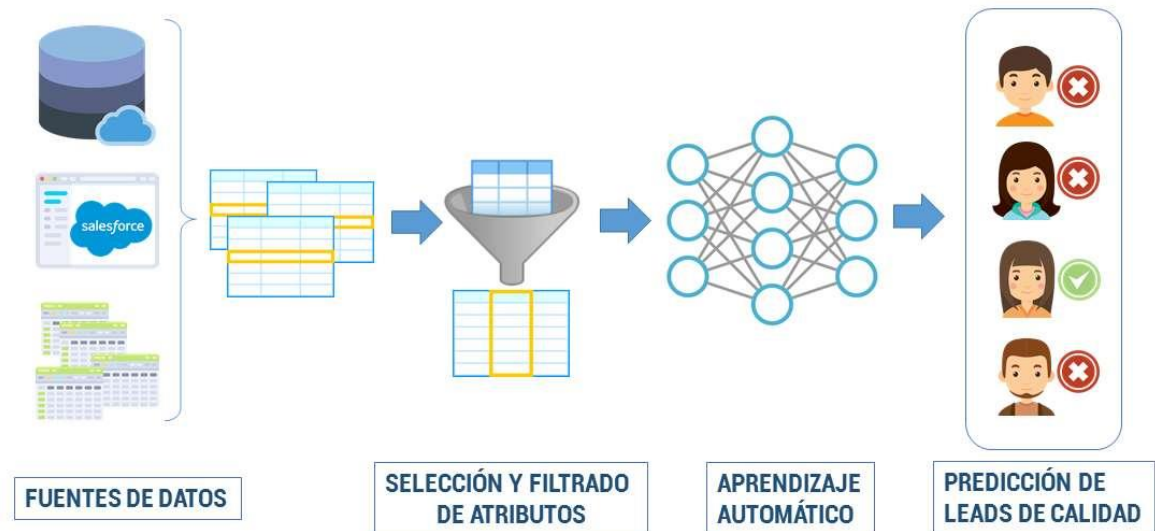


FUENTE: (Guerrero, 2018)

1.6.1.2 Minería de datos

La minería de datos es un proceso computacional que analiza volúmenes de información en la búsqueda de patrones y relaciones sistemáticas con un proceso que implementa métodos de aprendizaje, inteligencia artificial y estadística. Mediante la minería de datos se fundamenta en predecir una variable o un comportamiento siendo la principal razón de que se pueda implementar. Mediante diferentes softwares se pueden aplicar diversas transformaciones de la información, corregirla, resumirla, moldearla e incluso de proponer modelos para la solución de algún problema. (Berrios, 2014)

Ilustración 3. Proceso de minería de datos



FUENTE: (Puertas, 2016)

1.6.1.3 Modelos clasificatorios

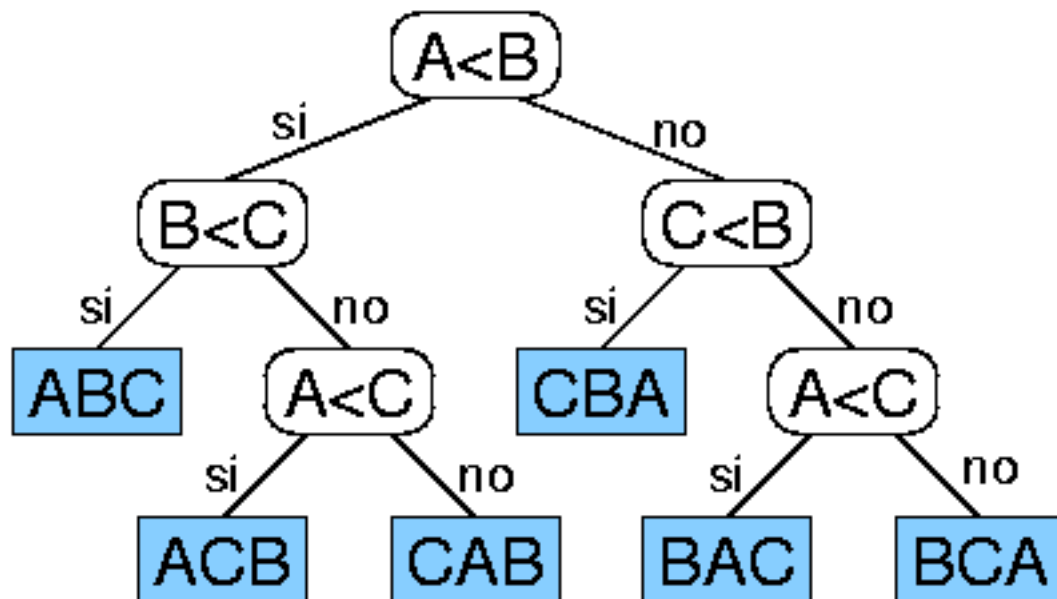
A lo largo de las investigaciones estadísticas van enfrentado el reto de manipular grandes cantidades de datos, los cuales incluyen un numero de variables ostentosas, con los cuales se debe tonar información, encontrar patrones y estructurar la información. Pero en 1971, Sonquist, Baker y Morgan propusieron el algoritmo de análisis de detección automática de interacciones o en sus siglas AID, el cual fue pionero de ajuste de datos, basados en arboles de decisiones. (Hadidi, 2003). Tiempo luego en 1980, Kass propone un algoritmo recursivo de clasificación llamado CHAID, finalmente aparecen unos métodos más recientes como lo son FIRM propuesto por Hawkins y MARS propuesto por Friedman y finalmente CART propuesto por Breiman, el cual se usa para la construcción de árboles de regresión y clasificación mediante algoritmos recursivos. (Hadidi, 2003).

- Árboles de decisión

Es una técnica de predicción y clasificación, la cual consiste en un desglose jerárquico y secuencial del problema a resolver, con el cual cada nodo representa gráficamente las decisiones posibles o los caminos que este pueda tomar con distintas combinaciones de decisiones y eventos. Estos árboles muestran distintas reglas que pueden ser entendidas por los humanos, ya que el conocimiento lo genera el mismo árbol y no debe tener conocimiento previo del tema. (Berrios, 2014).

Este algoritmo es empezado a construirse bajo un modelo jerárquico generalizado llamado Arriba-abajo, buscando de una manera de recursividad aquellos atributos que generen un árbol óptimo, con una estructura lo más simple posible. Los algoritmos más conocidos son el ID3, el C4.5, C5.0 y CART; Caracterizándose por tener la capacidad de procesar grandes volúmenes de información de una forma eficiente y error en los valores mediante el entrenamiento del mismo. (Quinlan, 1993).

Ilustración 4. Diagrama árbol de decisión



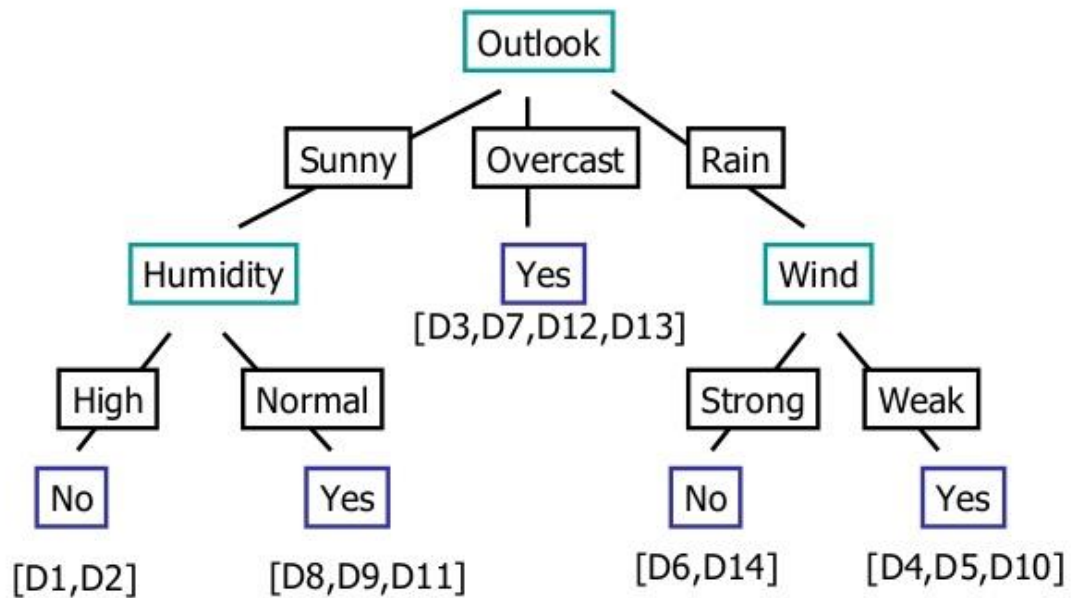
FUENTE: (Ciencias de la computación, s.f.)

- Algoritmo ID3

Fue un algoritmo propuesto en 1975 por J Ross Quinlan en su libro “Machine learning”, el cual construye un árbol de decisión con un conjunto de datos fijo, implementándolo para clasificar futuros ejemplos, mediante el cual cada ejemplo puede contener varios atributos que pertenecen a una clase es decir valores verdaderos o falsos. Cada nodo u hoja del árbol contienen los nombres de la clase, mientras que las ramas pertenecen a un posible valor del atributo. Este algoritmo se basa en la ganancia de información, el cual decide que atributo va en cada nodo, el cual manejando el concepto de entropía se puede aplicar la ganancia de información. (Quinlan, 1993).

Ilustración 5. Algoritmo ID3

Ejemplo: Algoritmo ID3



FUENTE: (Vidal, 2008)

- Algoritmo C4.5

Este algoritmo propuesto en 1993 por J Ross Quinlan, se denominó la versión avanzada del ID3, incluyendo manejo de valores continuos y discretos para manejar atributos continuos, manejo de valores de atributos faltantes usando ponderaciones de valores y probabilidades en lugar de valores cercanos o comunes, generar datos de reglas los cuales permiten interpretar de una forma más eficiente a cualquier árbol, finalmente este algoritmo construye un árbol simplificado y genera resultados de una manera más fácil de entender. (Quinlan, 1993).

Aunque este algoritmo es uno de los más usados tiene algunas desventajas como la presencia de ramas vacías, ramas poco significativas y finalmente sobreajuste cuando el algoritmo selecciona información con características que no son de real importancia, causando rompimiento en el proceso de distribución. (Berrios, 2014)

- Algoritmo C5.0

Este algoritmo construye árboles en base a un entrenamiento previo optimizado bajo la entropía implementando la ganancia de la información y se le denomina la evolución a su predecesor. Favoreciéndolo con la eficiencia del tiempo de construcción del árbol, el uso de memoria y la obtención de árboles más pequeños con la misma capacidad predictiva que su predecesor. (Berrios, 2014)

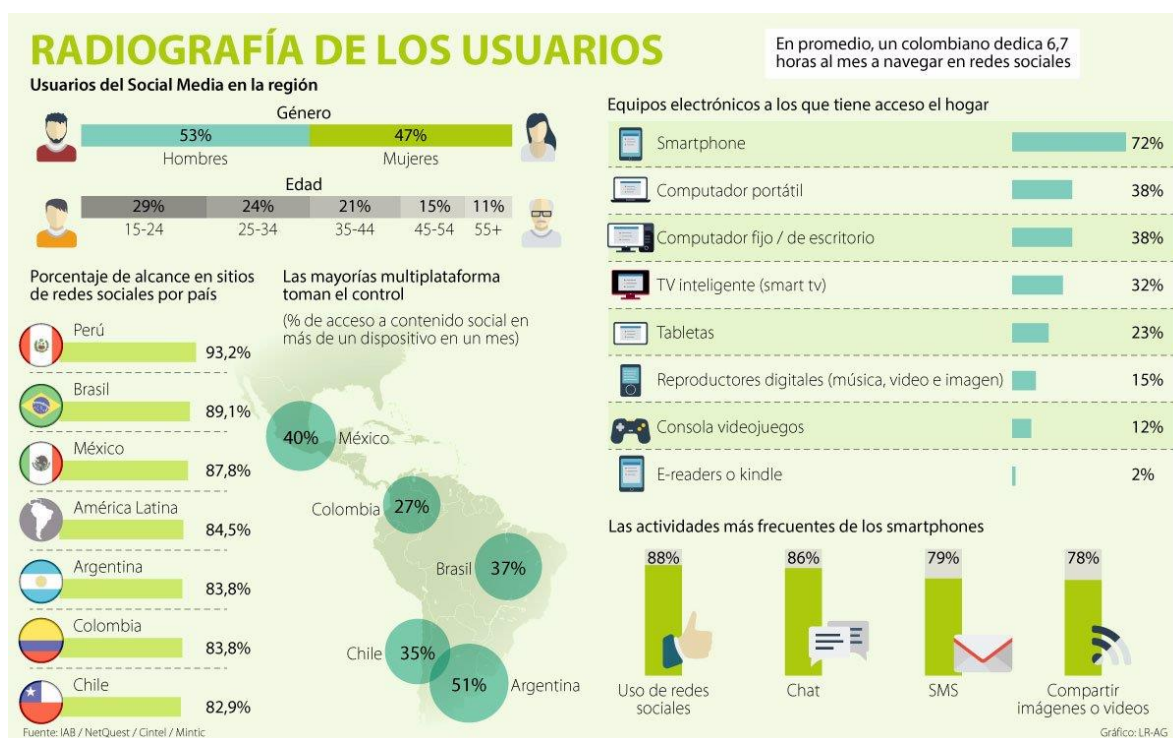
- CART: Arboles de clasificación y regresión

Este algoritmo desarrollado en 1984 por Breiman, cuyo resultado es un árbol de decisiones, es un método no paramétrico de segmentación binaria el cual se construye el árbol dividiendo repetidamente el set de datos. Su nodo inicial es llamado nodo raíz y se divide en hijos o nodos, posteriormente cada grupo de hijos es particionado por separados, seleccionando las divisiones de una forma que los errores sean menores que los del nodo raíz y definiéndolos por una variable explicativa, todo esto, con el fin de obtener un grupo homogéneo y mantener un árbol mucho más pequeños. (Zhang, Coomans, & Heyden, 2006).

1.6.1.4 Uso de redes sociales en Colombia.

Las redes sociales en Colombia son de gran impacto, ya que según el DANE nos encontramos en Colombia 45,5 millones de habitantes según el Censo Nacional de Población y Vivienda del DANE (Dinero, 2018). Por otro lado, las redes sociales generan datos a diario lo afirma la compañía ComScore, ya que realizó un estudio el cual identifica que un ciudadano colombiano puede gastar alrededor de 7 horas navegando en las redes sociales, ubicando a Colombia como el quinto país de la región con mayor uso de redes sociales, ubicándolo contra otros países como Perú, Brasil, Argentina y finalmente México, usando las redes sociales con una participación del 88%, acto seguido un 79% en chat y finalmente compartir imágenes en un 78%.²

Ilustración 6. Uso de redes sociales en Colombia



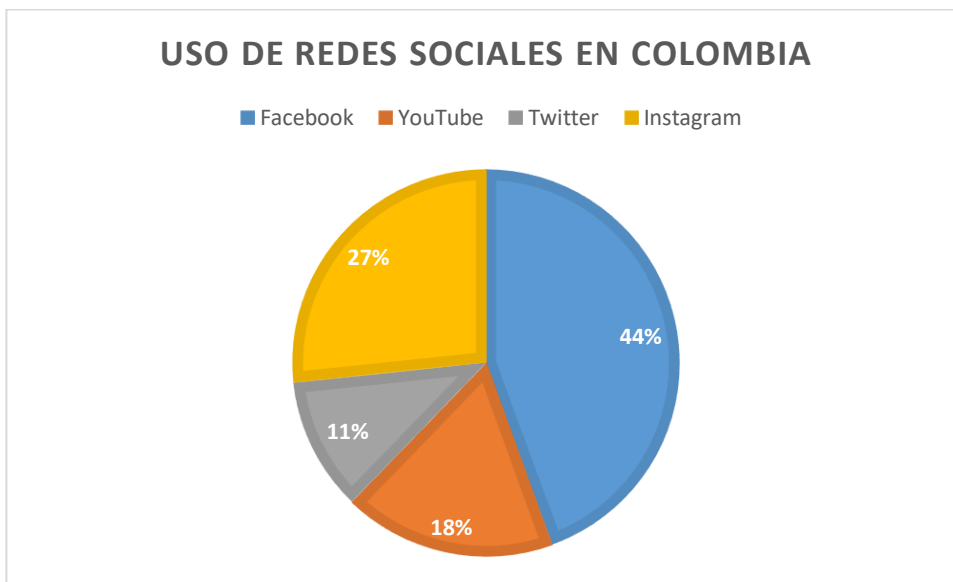
Fuente: (Cigüenza Riaño, 2019)

En la ilustración se puede apreciar como Colombia, es un país significativo para el tratamiento de información y procesamiento de la misma, y más aún en redes sociales debido a que la cantidad de información que se genera en las mismas es

² Disponible en La Republica <<https://www.larepublica.co/internet-economy/hasta-siete-horas-puede-gastar-un-colombiano-en-las-redes-sociales-2826041>>

de gran magnitud con distinto contenido. Finalmente, esta misma compañía realizo un estudio en donde identifico cuatro redes sociales importantes como lo son Facebook, Instagram, Twitter y YouTube, el cual Colombia ha tenido descenso en la audiencia de las mismas por teléfonos móviles, y en computador de escritorio es bastante baja, en la siguiente grafica se pueden identificar los usos de redes sociales en Colombia, identificando como Twitter una de las redes sociales que menos usuarios activos tiene. (Chiquiza Nonsoque, 2018).

Ilustración 7. Uso de Redes Sociales en Colombia



FUENTE: (AUTOR,2019)

1.1.1. Marco conceptual.

1.1.1.1. OWL.

Este lenguaje desarrollado e implementado por la web semántica, con el fin de representar conocimiento acerca de las cosas, grupo de cosas y todo lo relacionado entre ellas, para poder desarrollar ontologías y puedan ser procesadas por la computadora³. Por medio de OWL, el cual será el lenguaje que se utilizará para creación y edición de las ontologías permitirá el uso de la herramienta Protege y el uso estándar de RDF.

1.1.1.2. XML.

Es un estándar internacionalmente conocido para la creación de documentos y en la web semántica, provee una sintaxis en la estructura de los documentos, basado en texto, formado por datos de caracteres y marcado que forman las etiquetas. Los componentes de un XML son**:

- Elementos: son la pieza clave de una cadena de texto.
- Instrucciones: Ordenes especiales para ser utilizadas en la aplicación que lo procesa.

1.6.1.2 Ontology development 101

Es una guía creada con el fin de que las personas puedan realizar su primera ontología mediante pasos, la cual es desarrollada por la Universidad de Stanford con el fin de poder ser implementada en el software protégé, en esta guía se pueden ver ontologías básicas y comunes como la del vino y la comida, la cual sirven como ejemplos para la creación y adaptación de nuestras propias ontologías (Natalaya F & McGuiness, 2000). Cabe destacar que esta no es la única metodología y no está definida como tal para hacer ontologías, pero puede servir como referencia ya que es promovida por protégé.

1.6.1.3 Ionic

Es un framework que trabaja de la mano con Angular y nos puede ayudar para la creación de aplicaciones móviles, combinando estos dos, permiten crear aplicaciones híbridas ya sea en Android o IOS. Ionic maneja características importantes como lo son patrones de modelo, vista, controlador y modelo, vista,

³ Disponible en la W3C < <https://www.w3.org/OWL/>>

** Disponible en la web < <http://www.mundolinux.info/que-es-xml.htm>>

vista modelo, sistema de clases, control DOM, interfaz y temas, widgets, diseño adaptable, entre otras características que facilitan el desarrollo de las aplicaciones.⁴

1.6.1.4 Node.js

Node.js su lenguaje fundamental es JavaScript y este trabaja en el lado del servidor, permitiendo construir programas escalables utilizando programación orientada a eventos, permitiéndonos trabajar con API RESTful, tomando algunos parámetros, interpretándolos y finalmente armando y descargando una respuesta de vuelta al usuario.⁵

1.6.2.1 Protégé

Este framework open source suministra una interfaz gráfica para la creación y edición de ontologías, el cual es desarrollado por la Universidad de Stanford bajo el lenguaje de programación de java. Este marco de trabajo se decide implementar debido a su facilidad de uso, además de ser gratuito y cuenta con textos para su aprendizaje y uso de factores que influyan en la selección. (Cabrera & Guerrero, 2014)

1.6.2.2 Ontologías

Es un conjunto de palabras las cuales representan algún fin en específico, son parte importante de la web semántica, y son de vital importancia para introducir en los recursos un conocimiento estructurado, componiéndose de clases, propiedades e instancias las cuales forman las ontologías, describiendo y descomponiendo un dominio en específico (Thomas , 1993)

1.6.2.3 Twitter

Es clasificado como una red social en la que los usuarios interactúan entre ellos de manera inmediata en mensajes no más de 280 caracteres, como máximo, permitiendo a los usuarios comunicarse, buscar y seguir a la gente que más les llame la atención y sea del interés del mismo, a estos mensajes se le denominan tweets los cuales pueden ser vistos por cualquier otro usuarios que este registrado

⁴ JOSE PEREZ RIVAS. Hablemos un poco de ionic [en línea].

<<https://www.phonegapspain.com/que-es-y-como-empezar-con-ionic-framework/>> [citado el 24 de abril del 2019]

⁵ ABERNETHY MICHAEL. ¿Simplemente qué es Node.js? [en línea]. <

<https://www.ibm.com/developerworks/ssa/opensource/library/os-nodejs/index.html>> [citado el 27 de abril del 2019]

en la plataforma (solo si el usuario lo tiene disponible), ya que maneja una privacidad de cuentas, la cual el usuario puede modificarla y determinar quiénes pueden ver sus tweets. (Castelló Martínez, 2018)

2. DISEÑO METODOLÓGICO

Para comenzar en este trabajo investigativo se recopiló información en distintas revistas científicas y repositorios interinstitucionales, por consiguiente, se procedió a analizar y conllevar a la mejor solución del problema.

A continuación, se realizó el debido proceso investigativo y de desarrollo mediante la implementación de un modelo predictivo y aplicando como base un vocabulario ontológico, con el fin de lograr el cometido, llevando a cabo un trabajo exhaustivo y explicativo que llevaron a contemplar el logro satisfactorio de dicho planteamiento.

2 Metodología.

Este estudio obedece a un tipo de investigación exploratoria cuantitativa, debido a que se hace una recopilación de información y material para poder hacer un análisis acerca del problema a tratar, y de esta forma poder analizar, comparar y finalmente proponer soluciones.

La estructura de este trabajo investigativo es descriptiva, ya que se manipulan las variables, y se manipula el factor supuestamente causal, así como se aplican procedimientos para la selección y asignación de sujetos y tratamiento. Y finalmente es transversal descriptivo, porque se recopila información en un solo momento con el fin de describir variables y así poderlas analizar.

2.1 Método de recolección de datos.

En primera instancia, la inteligencia de negocios abarca un gran mar de conocimientos en los cuales puede usarse para resolver distintas problemáticas que se presentan día a día, del mismo modo distintos cuestionamientos y preguntas pueden llegarse a resolver mediante la inteligencia de negocios, por ende se tomó como alternativa un ETL que lograra ajustar a los requerimientos del modelo ontológico y poderlo mapear con el modelo predictivo, con el fin de lograr obtener un acierto en las posturas políticas de los usuarios en Twitter, pues bien aplicando este concepto se puede lograr a futuros proyectos que pueden abarcar no solo conceptos políticos, sino también conceptos comunitarios, financieros, educativos, entre otros.

En definitiva, mediante la implementación de sockets y la conexión streaming a la API de twitter, se obtienen los tweets en tiempo real, pasándolos por un proceso

ETL, un algoritmo predictivo y un entrenamiento del algoritmo con el vocabulario ontológico, se puede llevar acabo el prototipo que asemeje un favoritismo de hoy en día respecto al sentimiento de cada tweet.

2.3 Fuentes de información.

En primera medida las fuentes investigativas fueron secundarios de repositorios académicos y artículos sumamente científicos, los cuales mediante un exhaustivo análisis investigativo permitieron la culminación satisfactoria del proyecto con el fin de poder sustentar un modelo que permita la solución al problema.

CAPITULO 3

Método de extracción de datos enlazados

3. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 1 – DEFINIR EL MÉTODO PARA LA EXTRACCIÓN DE DATOS ENLAZADOS PARA ENCONTRAR LA INFORMACIÓN

En primera instancia como resultado esperado se definió para la extracción de la información una conexión por streaming al api de twitter, mediante la cual mediante peticiones REST una única conexión dando respuesta a todos los mensajes publicados en vivo vía twitter, acto seguido, el modelo en sus siglas ETL (Extract, transform, load), mediante el cual permite identificar todos aquellos patrones en la estructura de un twitter que no son relevantes y lograr centralizar los textos del tweet el cual es el foco de nuestro análisis. Posteriormente se tiene el análisis y diseño conceptual de la tendencia política del twittero, mediante la cual se implementará a cabo teniendo en cuenta la metodología *Ontology Development 101*, con el fin de poder identificar todos los elementos relevantes que se pretenden conceptualizar, y añadiendo la relación entre cada uno de ellos. Finalmente, y no menos importante se implementará la ontología con ayuda de la herramienta Protégé, teniéndola como una base sólida para el análisis y diseño realizado anteriormente, y no obstante se implementará esta ontología como datos de entrenamiento para nuestro modelo.

3.1 Análisis y diseño de la extracción, transformación y carga de datos.

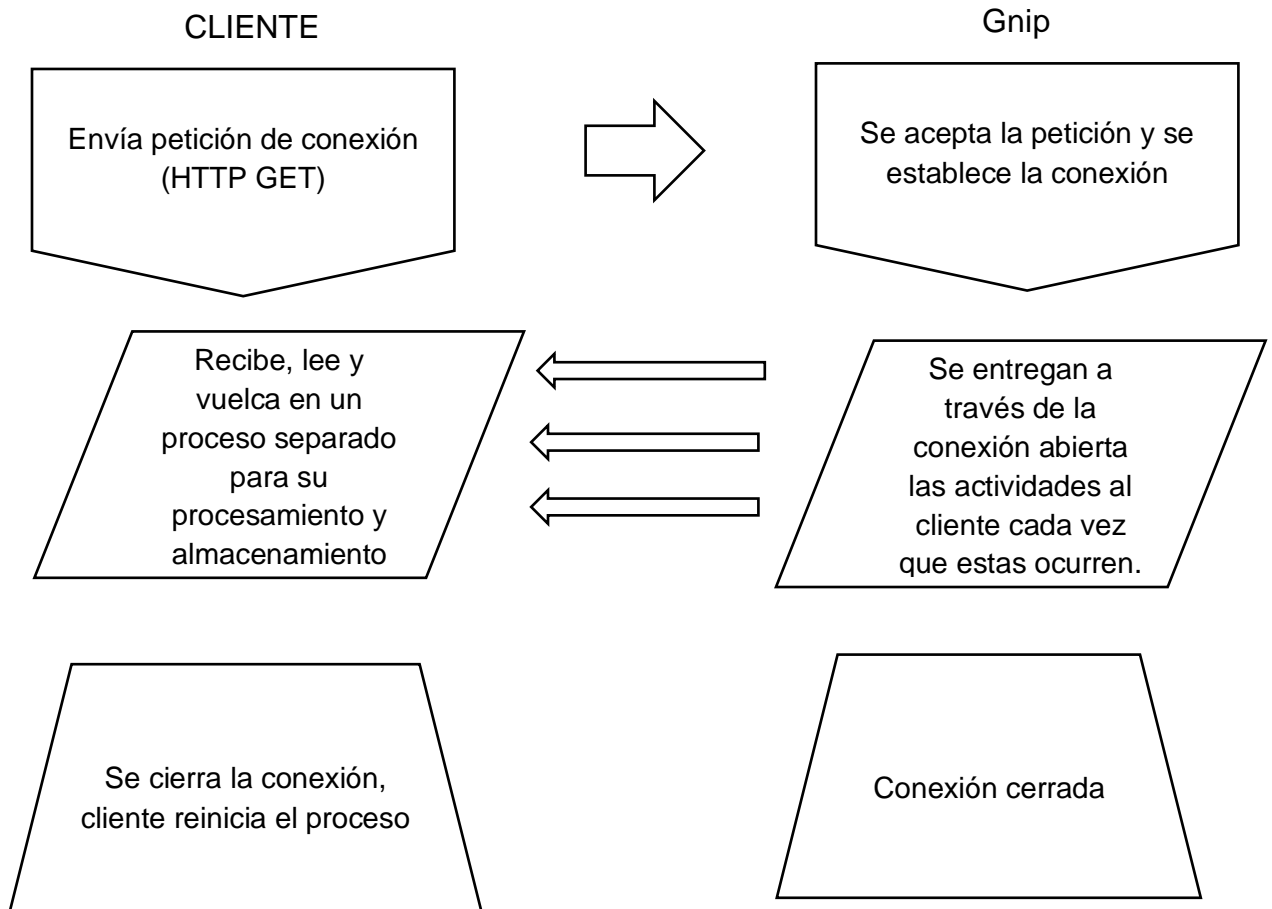
Para el desarrollo de este resultado esperado se implementó la extracción de datos vía streaming al api de twitter y se adoptó la extracción, transformación y carga de datos, con el fin de obtener los mensajes en tiempo real por los usuarios.

3.1.1 Extracción

Mediante un enlace por streaming al api de twitter, la cual permite por peticiones REST con una única conexión entre cliente y twitter, dando respuesta a todos los mensajes publicados en vivo vía twitter, la cual debe constar de una aprobación por la compañía suministrándonos unas llaves y fichas para poder conectarnos a esta. Nos devolverá como resultado la conexión aprobada, recibiendo toda clase de información que se esté generando en twitter. En la siguiente figura podemos

observar el proceso que se maneja para el enlace cliente y Gnip (api de twitter). ⁶

Ilustración 8. Conexión entre cliente y api twitter



FUENTE: (Autor ,2019)

3.1.3 Transformación

En esta fase de limpieza y transformación, se implementó un script desarrollado en Node.js, con el fin de filtrar toda la información que es de nuestro interés, por lo tanto, se implementó una función en el backend, con el fin de filtrar el tweet, lo cual es de suma importancia ya que este es el insumo para el tratamiento de la

⁶ TWITTER DEVELOPER. Filter realtime tweets [en línea]. <
<https://developer.twitter.com/en/docs.html>> [citado el 18 de abril del 2019]

información en nuestro modelo, en la siguiente figura se puede observar una parte del código que se implementó.

Ilustración 9. Script de filtro de información

```
var client = new Twitter(toker_twitter.api_twitter);  
var stream = client.stream('statuses/filter', { track:  
'@IvanDuque,@petrogustavo,@sergio_fajardo,@German_Vargas' });
```

FUENTE: (Autor,2019)

3.1.3 Carga

Para esta fase, los datos anteriormente transformados son enviados a nuestro algoritmo predictivo, con el fin de ir analizando en tiempo real cada mensaje relacionado con cada uno de los personajes presidenciales.

3.2 Modelo ontológico que representa tendencias políticas favorables y desfavorables.

Para esta etapa y mediante la metodología anteriormente mencionada *Ontology Development 101*, la cual permite identificar cada termino que se conceptualice en la ontología y poderlo relacionar entre ellos. La cual nos sugiere 7 procesos que se deben seguir para llegar a un resultado. (Natalaya F & McGuinness, 2000)

3.2.1 Determinar el dominio y el alcance de la ontología.

Según la metodología, se debe conceptualizar las siguientes preguntas para este primer proceso:

- ¿Cuál es el dominio que abarcará la ontología?

Serán todos aquellos temas de elecciones pasadas presidenciales en Colombia.

- ¿Para qué se utilizará la ontología?

Mediante esta ontología se entrenará el algoritmo, el cual analizara todos los tweets que lleguen en tiempo real y los podrá clasificar según el criterio “Positivo”, “Neutro” y “Negativo”, y se reflejaran estos resultados en una aplicación móvil, la cual permitirá al usuario dar información de la tendencia política en la que los tweets

tienen inclinación y cuál podría ser el próximo presidente de estado de acuerdo a los datos recibidos.

- ¿A qué tipo de preguntas la información en la ontología proveerá respuesta? Serán preguntas de carácter informativo acerca de las elecciones presidenciales colombianas, Como, por ejemplo:

¿Cuál es la tendencia política de los usuarios en twitter?

¿Qué tan favorable ven los usuarios de twitter a Iván Duque?

¿Qué tan favorable ven los usuarios de twitter a cualquier otro candidato a las elecciones presidenciales?

¿Existe alguna aplicación que me informe cual es la tendencia política en redes sociales, específicamente en twitter?

- ¿Quién usará y dará mantenimiento a esta ontología?

El director y autor de esta ontología dará el respectivo uso y mantenimiento de la misma.

3.2.2 Considerar reutilizar ontologías existentes

Es significativamente importante tener como referencia ontologías que abarquen temas del mismo a desarrollar, las cuales podríamos reutilizar o sirvan como una base sólida para crear una nueva **(Natalaya F & McGuiness, 2000)**. Por lo tanto, navegando en la web, específicamente en la biblioteca de ontologías como lo son Swoogle, DAML ontology library, Protége ontology library , Schema Web y finalmente GitHub, podemos evidenciar solamente dos ontologías con referencias política, que tuvieron como criterio de búsqueda campañas políticas, campañas financieras políticas, políticas, política colombiana, presidenciales, elecciones, congreso, gobierno, obteniendo de esta forma alrededor de 50 ontologías, las cuales sus links ya no existían, o simplemente no tenían contenido relacionado a política, debido a que las ontologías son creadas por usuarios y pueden ser llamadas a gusto del mismo, finalmente se obtuvieron dos ontologías que sobresalieron y son una base sólida para el desarrollo de la nuestra. **(d'Aquin & F. Noy, 2012)**.

- Politics Ontology: esta ontología es desarrollada por un usuario de la semanticweb.org , el cual en su perfil se denomina Jaime en el 2016, dicha ontología agrupa todos los términos referentes a política y genera esta

ontología, cabe aclarar, que su alcance es mucho mayor ya que esta ontología abarca conceptos políticos generales y no son tan específicos para nuestro proyecto y más aún nuestro panorama político, y la complejidad de la misma es alta y algunos conceptos no son de interés para el proyecto, como lo son demócratas, republicanos, sectores primarios, secundarios, terciarios y demás que no son referentes a la política colombiana.

- Presidential Campaign Finance Ontology: es una ontología desarrollada por Kamruzzaman Sarker, la cual fue implementada para evaluar las elecciones políticas y la influencia financiera para la presidencia de Donald Trump en el 2016, esta ontología fue la base para el desarrollo de nuestra ontología.

3.2.3 Enumerar términos importantes de la ontología

Según el modelo que se implementa, es necesario y de vital importancia elaborar una lista de términos relacionados al dominio de interés, mediante las siguientes preguntas. (Natalaya F & McGuinness, 2000).

- ¿Qué propiedades tienen estos términos?
- ¿Cuáles son los términos de los que se requiere hablar?
- ¿Qué nos gustaría decir acerca de dichos términos?

Realizando de esta forma el siguiente listado:

- | | |
|---------------|---------------------------|
| - Política | - Ciudadanía |
| - Disputa | - Participación ciudadana |
| - Propuestas | - Comunicación |
| - Tendencia | - Congreso |
| - Autonomía | - Beneficio |
| - Candidato | - Créditos |
| - Partido | - Cultura |
| - Región | - Agentes ambientales |
| - Populismo | - Generación eléctrica |
| - Agricultura | - Desarrollo |
| - Ejército | - Equidad |
| - Artes | - Sostenibilidad |
| - Sociedades | - Medio Ambiente |
| - Economía | - Salud |
| - Edad | |
| - Profesión | |

- Palabras agradables
- Palabras desagradables
- Nombre Candidato
- Educación
- Derechos Humanos
- Industria y tecnología
- Trabajo
- Financiación de campaña
- Ciudad
- Nombre Partido
- Ocupación
- Tendencia política

Este listado propuesto se debe gracias a las ontologías anteriormente mencionadas, las cuales fueron la base para el desarrollo, que posteriormente vamos a implementar para el siguiente paso.

3.2.4 Definir las clases y jerarquía de clases

En el desarrollo de la jerarquía de clases, se implementará el método top-down, el cual consiste en definir los conceptos más generales del dominio, para llegar al tope de los más especializados. En primera instancia, seleccionaremos aquellos términos que describen objetos reales y no dependen de otros; estos representaran las clases. (Natalaya F & McGuiness, 2000).

1. Candidato
2. Sociedades
3. Financiación de campaña

3.2.6 Definir las propiedades de las clases y definir los aspectos de las propiedades

Aquellos términos que no fueron incluidos en la lista anteriormente, son conocidos como propiedades de las clases, es decir brindan información de cada clase. (Natalaya F & McGuiness, 2000). Como por ejemplo “Ciudad”, “Ocupación”, “Tendencia política”, “Nombre partido”, hacen parte de las clases de primer nivel “Candidato”. Cabe aclarar, que pueden existir otras propiedades que surgen de otras relaciones entre clases, como por ejemplo “Propuestas” puede tener una propiedad, “pertenece al candidato”, que se relaciona a la clase “candidato”, o que “Sociedades” tiene una “Región” y puede relacionarse a la clase “Candidato”.

3.2.7 Crear instancia

Finalmente, debemos crear instancias de las clases de la jerarquía, mediante los siguientes pasos: (Natalaya F & McGuiness, 2000).

- Escoger una clase,

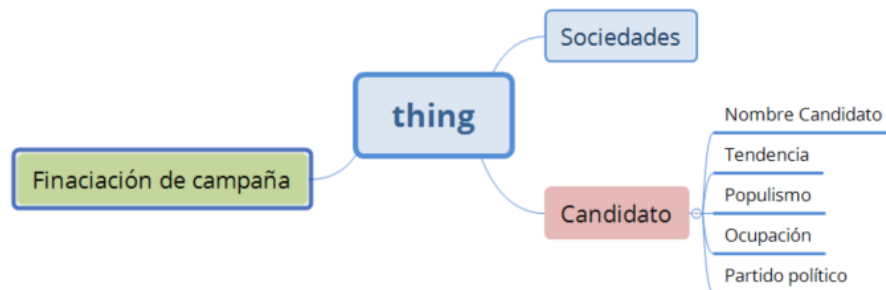
- Crear una instancia de esa clase.
- Llenar sus propiedades.

Como, por ejemplo, la clase “Nombre Candidato” y se instancio con Iván Duque, y llenando sus propiedades, “Nombre Partido”, “Ocupación”, “Tendencia Política” con la cadena de caracteres “Centro democrático”, “Abogado”, “Ultraderecha”, respectivamente.

3.2 Modelo ontológico que represente tendencias políticas.

Mediante el software Protége, desarrollado por la universidad de Stanford, se implementó la estructura de clases definida mediante el modelo anteriormente, desarrollado. En la ilustración 3 se puede visualizar la ontología desarrollada en Protége.

Ilustración 10. Ontología propuesta



FUENTE: (Autor, 2019)

Una vez realizado todos los pasos se presentan en la Tabla 1, todas las palabras que fueron anexadas en la ontología y que servirán como entrenamiento de nuestro algoritmo.

Tabla 1. Vocabulario ontológico

Candidato	@IvánDuque - @petrogustavo - @sergio_fajardo - @german_vargas
Tendencia	derecha, izquierda, centro
Clases sociales	proletariado, obreras, clase media, clase alta, clase baja
Ocupación	Abogado, político, matemático, economista
Partido político	Centro democrático, movimiento progresista, cambio radical, alianza verde
Propuestas	Agricultura, Ejército, Artes, Economía, Construcción, cultura, agentes ambientales, generación eléctrica, desarrollo, equidad, sostenibilidad, medio ambiente, salud, derechos humanos, industria y tecnología, empleo, educación, innovación, emprendimiento
Sociedades	Colombianas
Participación ciudadana	Democracia representativa, democracia, democracia inclusiva, interés general
Palabras agradables	Paz, Buen, excelente, tolerado, amado, representar, poder, riquezas, esfuerzo, buena remuneración, chévere, del putas, que chimba, chimba, abundar, abunda, abundancia, abundante, accesible, accesible, aclamación, aclamado, aclamación, espaldarazo, elogios, cómodo, acomodadito, realizar, consumado, logro, logros, preciso, precisamente, realizable, logro, logros, alcanzable, perspicacia, adaptable, adaptado, adecuado, ajustable, admirable, admirablemente, admiración, admirar, admirador, admirativo, admirativamente, adorable, adorar, adorado, adorador, cariñoso, adorando, hábil, hábilmente, adular, adulación, adulador, avanzado, ventaja, ventajoso, ventajosamente, ventajas, aventurero, aventurero, abogado, defendido, defensores, afabilidad, afable, afablemente, afectación, afecto, cariñoso, afinidad, afirmar, afirmación, afirmativo, afluencia, afluente, permitirse, asequible, asequible, asequible, ágil, ágilmente, agilidad, agradable, amabilidad, agradablemente, todo al rededor, seductor, seductoramente, altruista, altruista, asombro, asombrado, asombro, asombro, increíble, espantosamente, ambicioso, ambiciosamente, mejorar, dócil, amenidad, amabilidad, maliciosamente, amable, amabilidad, amistoso, amigablemente, amistad, amplio, ampliamente, entretener, divertido, graciosamente, negociar, capaz, beneficioso, mejor, mejor conocido, mejor de lo esperado, inocente, felicidad, feliz, felizmente, brillante, brillantez, fraternal, enérgico, calma, capaz, celebrar, cautivar, cautivador, elegante,

	claro, transparente, inteligente, hábilmente, compromiso, avanzar, felicitar, felicidades, constructivo, mejor país, valor, valiente, cortes, pacto, atrevidamente, bueno, decisivo, arriesgado, con carácter, sin tapujos, dedicado, orgulloso, derrotando, estupefacto, serio
Palabras desagradables	Terrorista, paraco, corrupto, inepto, traicionero, doble, arrodillado, conformista, títere, guerrillero, paramilitar, subversivo, bombero, cagan, caramelear, odiado, chanchullo, poco esfuerzo, chanfa, chichipato, chuzadas, conflicto armado, tirofijo, paila, guerra, mal, nada interesante, rabia, maluca, oportunista, incompetente, raro, no pude, necesita de, odio, aburrido, enfadado, enfado, culpa, inútil, bueno para nada, fracaso, fracasado, malo, perverso, desadaptado, incompresible, bueno para nada, manejado, manipulado, 2 caras, detener, frenar, estancar, limpieza social, águilas negras, matón, falsos positivos, cazador, incompetente, anormal, abolir, abominable, abominar, abominación, abortar, abortar, abortado, desgastar, abrasivo, abrupto, abruptamente, fugarse, ausencia, despistado, ausente, absurdo, absurdo, absurdamente, absurdo, abuso, abusado, abusos, abusivo, abismal, abismalmente, abismo, accidental, abordar, maldito, acusación, acusaciones, acusar, acusa, acusando, acusatoriamente, exacerbar, acerbo, acerbicamente, dolor, dolido, dolores, dolor, acre, acriticamente, acritud, mordaz, agravación, agresión, agresivo, agresividad, agresor, ofender, agonizar, agonizante, agonizante, agoní, afligir, enfermo, enfermedad, sin objetivo, alarma, alarmado, alarmante, alarmantemente, enajenar, alienado, alienación, alegación, alegatos, alegar, alérgico, alergias, alergia, adistancia, altercado, ambigüedad, ambiguo, ambivalencia, ambivalente, emboscada, amputar, anarquismo, anarquista, anarquista, anarquía, anémico, enfado, furiosamente, enojo, enojado, angustia, animosidad, aniquilar, aniquilación, molestar, molestia, molestias, irritado, molesto, molesto, molestar, anómalo, anomalía, antagonismo, antagonista, antagonista, desconcertado, desconcierto, desconcertante, cebo, obstáculo, banal, vulgarizar, perdición, desterrar, destierro, arruinado, bárbaro, bárbaro, bárbara, barbarie, bárbaro, bárbaramente, estéril, infundado, golpetazo, golpeó, tímido, paliza, Bastardo, bastardos, abollado, cañoneo, chalado, osuno, bestial, algarabía, ensuciar, mendigar, mendigo, miserable, mendicidad, engañar, apalea, tardí, desmentir, empequeñecer, menospreciado, menosprecio, belicoso, beligerancia, beligerante, beligerante, lamentar, lamentando, perplejo, doblado, regañarse, privar, duelo, privado, enloquecido, implorar, acosar, sitiar, manchar, traicionar, traición, traiciones, traidor, traicionando, traiciona, llorar, tener cuidado peor país, fraudulento, tirano, confundir, desconcertado, desconcertante, desconcertante mente, desconcierto, hechizar, parcialidad, parcial, sesgos, disputa, f

	anatismos, fanatismo, perra, malévolo, mordaz, mordazmente, amargo, amargamente, amargura, extraño, chismoso, chantaje, paja, culpa, culpable, amable, ablandar, blasfemar, blasfemo, blasfemia, maldito, evidente, descaradamente, decir tonterías, escupitajo, insolente, perverso, sin sentido común, manejado por otros, no, desolado, desoladamente, desolación, sangrar, sangría, sangra, defecto, ciego, cegador, ampolla, abrasador, hinchado, bloqueo, estúpido, matanza, sanguinario, sangriento, enrojecido, soplo, torpeza, torpe, errores, embotar, difuminar, borroso, borrón, borroso, difumina, decir bruscamente, no tiene pantalones, falta carácter, sobresaltarse, falso, hervir, hirviendo, bullicioso, bomba, bombardear, bombardeo, esclavitud, loco, aburrir, aburrido, aburrimiento, aburrimientos, aburrido, salpicón, voltearepismo
Región	Amazonia, andina, caribe, insular, Orinoquia, pacífico
Financiación de campaña	Partido político, canales privados, empresas públicas, empresas privadas, ingresos de votos
Gobierno	Congresistas, curules, presidente de la república, vicepresidente de la república, ministros, ministra del interior, ministro de relaciones exteriores, ministro de hacienda, departamentos administrativos, consejeros
Autonomía	Álvaro Uribe, Gustavo Bolívar, Antanas Mockus, Holman Morris, Claudia López,
Controversia	Timochenko, Manuel Marulanda, farcpolítica, farc, uribista, uribismo, guerrilla

Para poder obtener un adecuado árbol y de esta forma obtener las palabras adecuadas se implementó un diccionario de la terminología política colombiana⁷ y un lexicón de sentimientos que sirve como base para algunas palabras en cuanto a los sentimientos que se dan a las personas.⁸

⁷ BOCANEGRA VARÓN, Alfredo. Diccionario de la terminología política colombiana. Bogotá, 2015, 102 h. Tesis de maestría (Maestría en ciencia política)- Opinion lexicon

⁸ LIU BING, Sentimiento Análisis [en línea]. < <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar> > [citado el 25 de abril de 2019]

CAPITULO 4

Modelo predictivo para el procesamiento de datos enlazados

4. RESULTADOS ESPERADOS DEL OBJETIVO ESPECIFICO 2 – ESTABLECER UN MODELO PREDICTIVO QUE EVIDENCIE UN RESULTADO CERCANO A LA REALIDAD DE LOS DATOS EXTRAIDOS.

En este capítulo, se define el modelo predictivo a trabajar, con el fin de poder obtener el camino correcto para trabajar con el vocabulario ontológico previamente desarrollado, y poder entrenar el algoritmo en función de un análisis de sentimientos correcto. Para el modelo de análisis predictivo se seleccionó un algoritmo que permitiera emplear tendencias y patrones de comportamiento mediante ontologías, identificando relaciones entre patrones y variables que permitan llegar a un resultado, previo a una minería de datos. **(Fontes & Daradoumis, 2017).**

4.1 Modelos predictivos que aplican para vocabularios ontológicos.

En primer lugar, para la selección del modelo predictivo se tuvieron en cuenta puntos como la búsqueda de patrones de datos ocultos para responder datos que se relacionen unos a otros, como aplica para este trabajo de grado, de acuerdo al sentimiento que el usuario trata de inducir mediante sus 280 caracteres en un tweet o re tweet, de igual forma, consolidando los resultados de forma gráfica en el prototipo , con el fin de poder evaluar un porcentaje acertado y así optar por un posible resultado, tratando de inducir a su predicción de la forma correcta.

Las técnicas aplicables para el análisis predictivo pueden agruparse en técnicas de regresión véase la tabla 1 y técnicas de aprendizaje computacional véase la tabla 2, cabe aclarar que en las siguientes tablas no son todas las técnicas que se pueden aplicar, debido a que existen otras más pero no son objeto para el estudio del presente escrito. **(Fontes & Daradoumis, 2017).**

Tabla 2. Técnicas de regresión

Técnica	Consiste en
Modelo de regresión lineal	Este modelo analiza aquellas relaciones entre una variable dependiente y un conjunto de variables independientes
Análisis de supervivencia o duración	Este modelo emplea una tasa de riesgo, lo cual hace parte de que el evento pueda ocurrir en un tiempo condicional
Análisis de clasificación y regresión	Este modelo es una técnica de aprendizaje de árboles de decisiones, los cuales desglosan de la mejor forma observaciones basadas en una variable dependiente, y se empieza a dividir entre nodos hasta que tenga un final
Curvas de regresión adaptativa multivariable	Este modelo construye unos tipos de modelos flexibles en el cual un nodo pasa de un modelo de regresión local a otro y crea el punto de intersección entre estas dos curvas, dando como resultado la decisión.

FUENTE: (Fontes & Daradoumis, 2017).

Tabla 3. Técnicas de aprendizaje automático

Técnica	Consiste en
Redes neuronales	Estas son técnicas de modelado no lineal las cuales permiten modelas distintos tipos de funciones complejas.
Máquinas de vectores de soporte	Este tipo de técnica se emplea para detectar y explotar patrones complejos de datos, agrupándolos, ordenándolos y clasificándolos.
Näive Bayes	Este modelo se fundamente en la probabilidad condicional de Bayes
K-vecinos más cercanos	Este conjunto de entrenamientos con valores positivos y negativos, clasifica y calcula la distancia entre el vecino más cercano hasta clasificar la muestra.

FUENTE: (Fontes & Daradoumis, 2017)

4.2 Modelo predictivo para ontologías mediante el algoritmo de árbol de decisiones.

De acuerdo a las técnicas mencionadas con anterioridad y tomando como referencia artículos científicos como “Ontology-Based Decision Tree Algorithm for recommender systems” (Amancio & Reif, 2017), “Ontology enhancement through inductive decision tree” (Amancio & Reif, 2017), “Decision Tree Applied to learning relations between ontologies” (Mirambicka R & Sulthana A) , los cuales son los pilares para el desarrollo el presente proyecto y tomo como referencia el entrenamiento de los algoritmos mediante las ontologías implantadas en los artículos anteriormente mencionados y sumando a esto los artículos “A linked data-based decision tree classifier to review movies” (Aldarra & Muñoz, 2015), “Development of prediction model for linked data based on the decisión tree – for track a, task a1” (Dongkyou & Kim, 2019), en los cuales tomaron tecnologías como lo son linked data las cuales emplean ontologías para el adecuado desarrollo de sus datos de entrenamiento, se recurre a implementar el algoritmo de árboles de decisión, con el fin de poder disminuir incertidumbres, ya que en ningún caso se puede predecir lo que va a suceder, pero con un vocabulario ontológico correcto y un algoritmo con un sentido de entrenamiento adecuado, puede aspirar a encontrar

ciertos patrones que orienten una toma de decisiones asertiva, para este caso el análisis adecuado del tweet que el usuario ha realizado.

Una vez realizado el vocabulario ontológico, el cual servirá como insumo para el entrenamiento del algoritmo y procesada la información podrá tomar decisiones una vez son recibidos los tweets.

Para la implementación del algoritmo se construyó bajo Node.js y una función recursiva la cual fue implementada por una librería que nos provee npm llamada "Node decisión tree for nodejs" la cual es entrenada por el vocabulario ontológico y posteriormente recibe los tweets.

Una vez los tweets son pasados por nuestro algoritmo, este los empieza a clasificar según el sentimiento del usuario ya sea "Positivo", "Neutro" o "Negativo" y estos son emitidos mediante sockets y la librería socket.io al frontend el cual se encargara de escuchar lo que el backend transmite y estos puedan ser representados en nuestro aplicativo móvil, como lo podemos evidenciar en el fragmento de código tomado del back, el cual es la parte final de nuestro desarrollo, contemplando los resultados y almacenándolos en una base de datos no relacional.

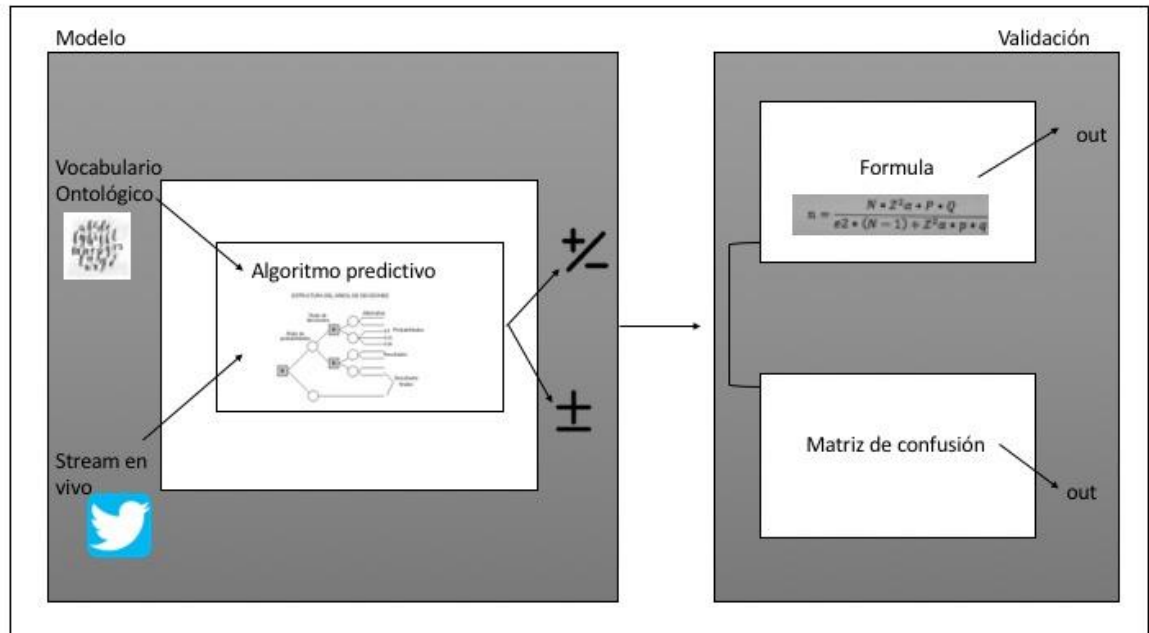
Ilustración 11. Script de análisis mediante el árbol de decisión

```
let analisis = {  
  "@IvanDuque": {  
    positivo: 0,  
    negativo: 0,  
    neutro: 0  
  },  
  "@petrogustavo": {  
    positivo: 0,  
    negativo: 0,  
    neutro: 0  
  },  
  "@sergio_fajardo": {  
    positivo: 0,  
    negativo: 0,  
    neutro: 0  
  },  
  "@German_Vargas": {  
    positivo: 0,  
    negativo: 0,  
    neutro: 0  
  }  
}
```

FUENTE: (Autor ,2019)

El algoritmo de árbol de decisiones debe contemplar el vocabulario ontológico como base para su entrenamiento con el fin de poder procesar los datos obtenidos por sockets y poderlos clasificar según correspondan, este algoritmo permite que se le puedan añadir más palabras al vocabulario permitiendo entrenar cada vez mejor y con mejores resultados.

Ilustración 12. Modelo predictivo



FUENTE: (Autor ,2019)

En la ilustración 10, se puede observar el modelo implementado a partir del vocabulario ontológico y el set de datos de entrenamiento para el algoritmo predictivo, el cual permite obtener el análisis de los estados de las personas en sus respectivos mensajes.

Finalmente, la evaluación de los datos y el algoritmo mediante la implementación de la muestra finita y la matriz de confusión como precisión del algoritmo implementado, dan como resultado un modelo predictivo acorde al problema a resolver.

CAPITULO 5

Arquitectura de software para el desarrollo del modelo

5. Resultados esperados del objetivo 3 – diseñar una arquitectura de software que cumpla con requisitos funcionales y no funcionales para la implementación del modelo.

Ya que existen diferentes componentes, los cuales son implementados para llevar a cabo la publicación como el mismo consumo de información, debemos identificarlos plenamente y determinar su interacción.

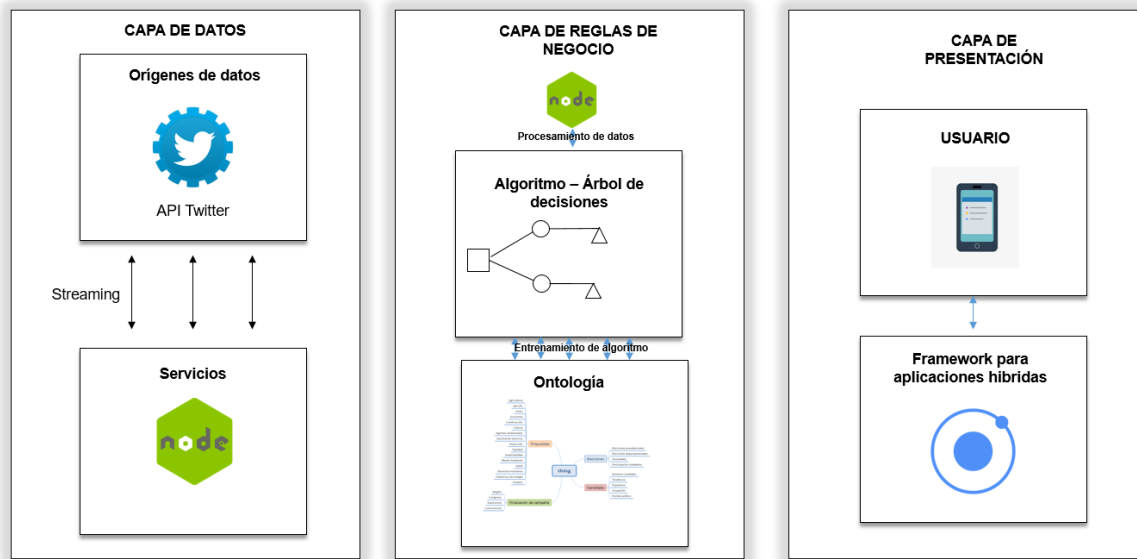
5.1 Arquitectura de software.

En la ilustración siguiente podemos evidenciar todos aquellos elementos que se relacionan y están involucrados en la aplicación. Cabe destacar, que el vocabulario ontológico desarrollada “predictionsColombia”, es mapeada al árbol de decisiones, dando como resultado una matriz que servirá como entrenamiento al algoritmo para que pueda ser recorrida con una función recursiva y de esta forma obtener el resultado adecuado. Con esta aplicación móvil, con nombre JereSoft accederá, para poder hacer consulta de los resultados obtenidos que esta función devolverá en un objeto.

- Capa de datos: esta capa tendrá la API de Twitter como origen de datos, con el fin de poder procesar la información mediante el ETL, obteniendo vía streaming los tweets de los usuarios y filtrándolos de acuerdo a las reglas establecidas, como servidor se tiene a la herramienta Node.js para el adecuado procesamiento.
- Capa de reglas de negocio: con el fin de procesar los datos, la información que se genera una vez procesada, es conectada al algoritmo el cual contiene el vocabulario ontológico como base del entrenamiento del mismo, este algoritmo responde y envía los datos mediante Node.js implementando sockets con el fin de poder ser visualizados en la capa de presentación.
- Capa de presentación: esta capa contiene el framework implementado para aplicaciones híbridas, en este caso Ionic con el fin de obtener el ultimo resultado y este será el usuario con la correcta visualización de los resultados obtenidos desde cualquier dispositivo.

Ilustración 13. Diagrama general de los componentes involucrados para la solución

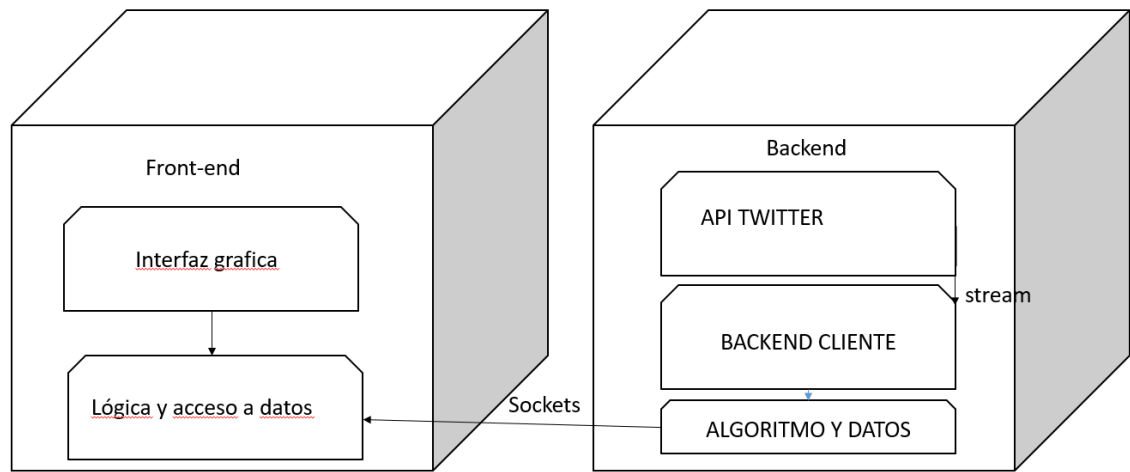
propuesta.



FUENTE: (Autor, 2019)

En la ilustración siguiente podemos apreciar la vista de despliegue del aplicativo móvil, los componentes de la interfaz gráfica y su lógica de acceso a los datos, la cual estarán instalados localmente por el framework Ionic.

Ilustración 14. Vista de despliegue de la aplicación móvil



FUENTE: (Autor, 2019)

Finalmente, se puede apreciar que se trabaja con una arquitectura modelo, vista, controlador, permitiendo trabajar en el backend con bases de datos no relacionales y en el frontend con vistas bajo el framework implementado.

CAPITULO 6

Prototipo móvil

6 Resultados esperados del objetivo específico 4 – desarrollar una prueba de concepto, basada en la arquitectura establecida.

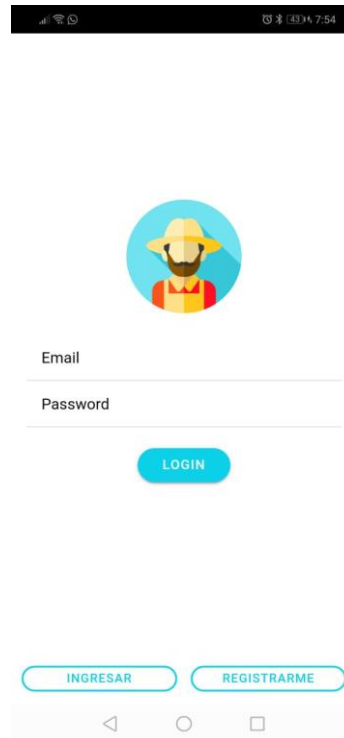
Una vez finalizado todo el proceso investigativo y los scripts necesarios para la prueba de concepto del funcionamiento adecuado del algoritmo implementado, se realiza una POC, la cual permitirá una experiencia de usuario agradable a los resultados generados en tiempo real. Lo que se busca mediante este objetivo es implementar una POC funcional de un aplicativo móvil, que conectada streaming a la API de twitter, procese la información, sea filtrada, logre llegar al algoritmo previamente entrenado e informe acerca de los favoritismos electorales para poder concluir un próximo presidente de estado.

6.1 Prototipo de aplicación móvil.

Finalmente se tiene como resultado esperado, el desarrollo de una aplicación móvil llamada JereSoft, la cual se escuchará por sockets la información procesada en el backend y podrá ser presentada al usuario con el fin de que pueda obtener la información de forma inmediata y en cualquier circunstancia, debido a que es una aplicación móvil se empleó un framework llamado Ionic con el fin de poder implementar la aplicación de forma híbrida y ser consultada en los dispositivos.

En la ilustración 15, se presenta la pantalla de autenticación, en la cual los usuarios deben introducir sus credenciales con los cuales se han registrado, una vez han introducido sus credenciales y son satisfactorios el servidor devolverá un token implementado por JWT, el cual será el acceso único para el usuario.

Ilustración 15. Pantalla de autenticación



FUENTE: (Autor, 2019)

En la ilustración 16, presenta la pantalla de registro, la cual si el usuario no cuenta aún con credenciales podrá registrarse enviando de esta forma al servidor una petición y se almacenará en una base de datos no relacional específicamente en MongoDB, las credenciales del usuario, para mayor seguridad en la aplicación, la clave del usuario será encriptada y de esta forma brindará seguridad en el aplicativo.

Ilustración 16. Pantalla de registro

Seleccione Avatar

Email

Nombre

Password

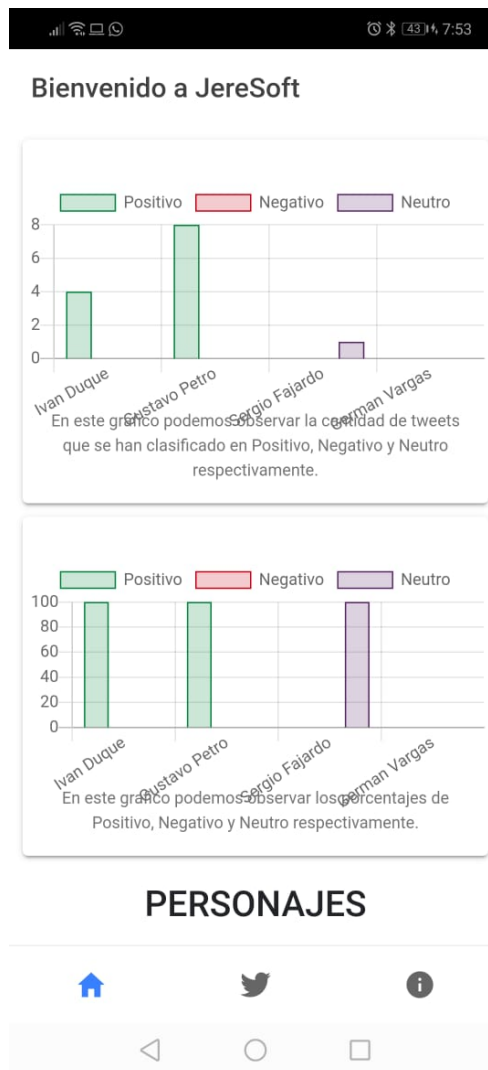
CREAR USUARIO

INGRESAR REGISTRARME

FUENTE: (Autor, 2019)

En la ilustración 17, presenta la pantalla del dashboard, la cual muestra dos tipos de graficas una con la cantidad de tweets que lleva un candidato y otra con los porcentajes de tweets que se estos van teniendo respectivamente, cabe denotar que estas graficas se van actualizando cada vez que algún usuario realiza un tweet respecto a los filtros establecidos en el capítulo 3. Finalmente, y no menos importante en la parte inferior podemos encontrar los personajes con los cuales se desarrolló esta POC y un breve resumen acerca de ellos extraído de Wikipedia.

Ilustración 17. Dashboard



FUENTE: (Autor, 2019)

En la ilustración 18, presenta la pantalla de los tweets en tiempo real, en esta podemos comprar todos los tweets que se están escribiendo en ese preciso momento, están organizados por el ultimo tweet puesto, hasta el más actual de forma ascendente, esta pantalla se implementa para poder verificar algún tweet en específico que queramos observar o ver los tweets que se van analizando en ese momento, podemos ver el nombre del usuario que realiza el tweet y el tweet correspondiente.

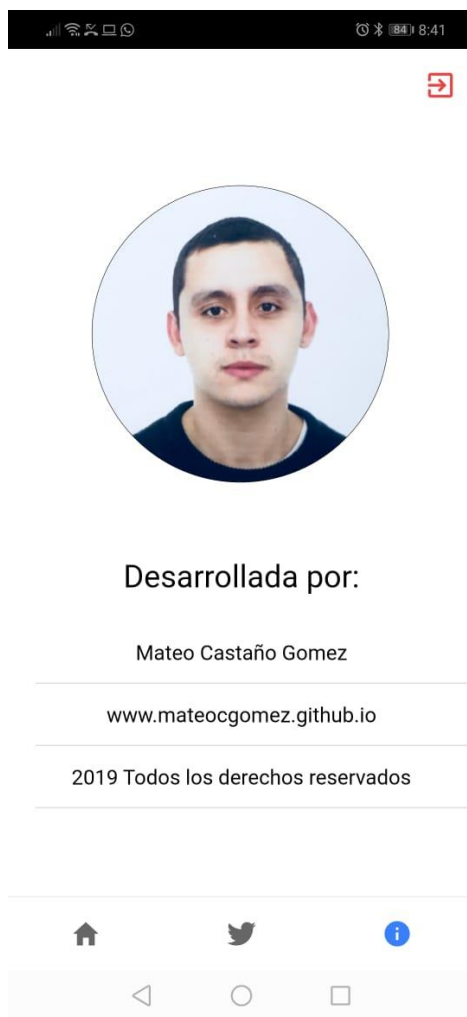
Ilustración 18. Tweets en tiempo real



FUENTE: (Autor, 2019)

En la ilustración 19, presenta la pantalla de información, esta muestra el autor de la aplicación y su contacto.

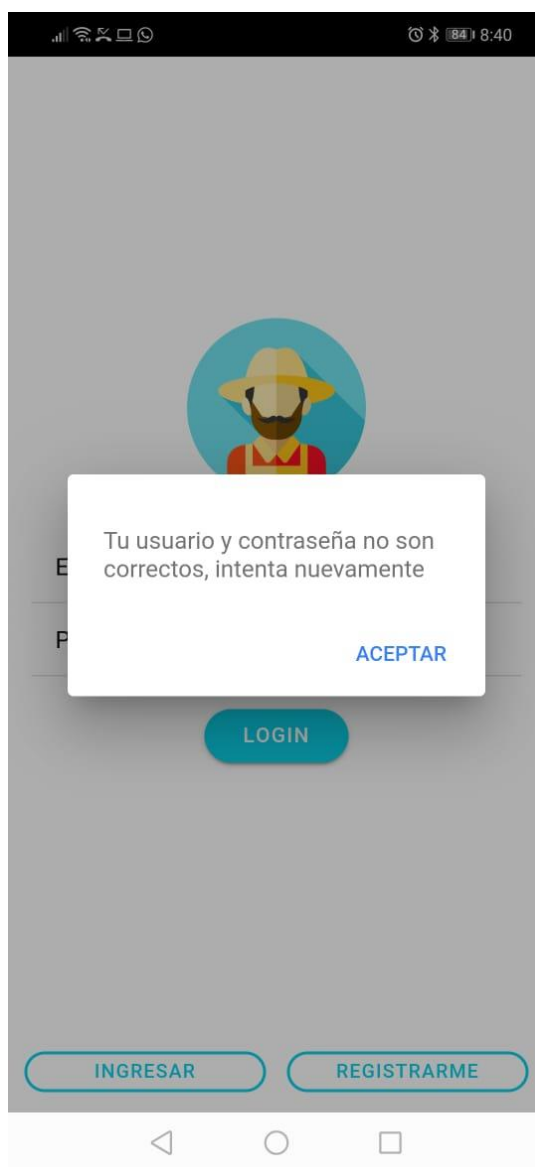
Ilustración 19. Información del autor



FUENTE: (Autor, 2019)

En la ilustración 20, presenta la pantalla de alertas, esta muestra la alerta que se contempla en la aplicación y fue desarrollada como un componente para poder ser implementada a futuro para otro tipo de anuncios.

Ilustración 20. Alertas



FUENTE: (Autor, 2019)

7. ANÁLISIS DE RESULTADOS

Una vez realizado la prueba de concepto, se debe establecer la matriz de confusión como una herramienta para evaluar el desempeño del algoritmo, la cual se implementa como análisis de resultados para evaluar la ontología y el algoritmo correctamente entrenado.

Para la ejecución de la evaluación se toma la muestra de una población con el fin de obtener la cantidad de tweets que se deben analizar manualmente para poderlos comparar con el modelo, los cuales serán evaluados mediante el algoritmo y por otra parte mediante intervención manual, estos datos clasificados manualmente se basarán en la ontología y el árbol descrito.

7.1 Calculo de muestra finita.

El desarrollo de la muestra finita, se debe implementar para calcular el valor de tweets que debemos analizar, debido a que la API de Twitter no es paga y solamente nos permite obtener 3200 tweets, por lo tanto, realizamos una muestra finita para tomar la muestra adecuada.

Ecuación 1. Formula de muestra

$$n = \frac{N * Z^2 \alpha * P * Q}{e^2 * (N - 1) + Z^2 \alpha * p * q}$$

En donde:

Parámetro	Insertar valor
N (tamaño de la muestra)	3200
Z (Nivel de confianza 50%)	0.674
P (variabilidad negativa)	50%
Q (variabilidad positiva)	50%
E (error)	3%

Obteniendo como resultado un total de 121 tweets, los cuales nos podrán indicar un 50% de veracidad en la implementación de nuestro modelo.

7.2 Resultados obtenidos por el modelo – datos de entrenamiento.

La POC se debe ejecutar en el transcurso de un tiempo determinado, permitiendo trabajar el algoritmo y obteniendo los 121 tweets que son implementados como base, obteniendo los resultados en la tabla siguiente:

Tabla 4. Resultados obtenidos mediante el árbol de decisión

	Positivo	Negativo	Neutro
Total	71	38	12

Fuente: (Autor, 2019)

7.3 Resultados mecánicos – datos de test

Al tomarse los 121 tweets, se clasifican de forma manual de acuerdo a la ontología con el fin de poder tener los datos de test para poderlos evaluar en la matriz de confusión.

Tabla 5. Resultados obtenidos manualmente

	Positivo	Negativo	Neutro
Total	69	39	13

Fuente: (Autor, 2019)

7.4 Matriz de confusión

La herramienta de matriz de confusión evalúa en las columnas las observaciones y en las filas las predicciones, para este caso se considera una clasificación binaria. Como primera medida se encuentran los “VP” los cuales son todos los positivos clasificados correctamente, en la segunda medida se encuentran los “FP” los cuales son los clasificados incorrectamente como positivos, en tercera medida se encuentran los “VN” los cuales son todos los negativos que fueron clasificados correctamente como negativos, como cuarta medida se encuentran los “FN”, los cuales son todos los positivos que fueron clasificados incorrectamente como negativos, como quinta medida tenemos los “MP” los cuales son todos los neutros clasificados correctamente, finalmente tenemos los “MN” los cuales son todos los neutros clasificados incorrectamente.

Tabla 6. Matriz de confusión

	Positivos	Negativos	Neutros	Predicción total
Positivos	52	6	13	71
Negativos	10	28	0	38
Neutros	7	5	0	12
Total	69	39	13	

Fuente: (Autor, 2019)

Una vez tenemos la matriz de confusión implementada se procede a realizar los cálculos respectivos.

Mediante la siguiente formula se obtiene el dato el cual el modelo clasifica correctamente.

Ecuación 2. Exactitud del algoritmo

$$A = \frac{(VP + VN + MP)}{TOTAL}$$

$$A = \frac{(52 + 28 + 0)}{(52 + 6 + 13 + 10 + 28 + 0 + 7 + 5 + 0)}$$

El resultado obtenido es de 0,66 es decir de un 66% de exactitud.

8. CONCLUSIONES

El desarrollo de este trabajo permitió analizar los sentimientos de los usuarios de Twitter, validando el método propuesto con el cual se obtuvo un 66% en exactitud del mismo, en donde se obtuvieron tweets de todo índole político, en el cual se procesó la información en tiempo real, permitiendo que el mismo filtrara y analizara los tweets permitiendo consolidar los sentimientos políticos de varias publicaciones realizadas por los personajes políticos y la reacción de los usuarios que aplican en esta red social, permitiendo analizar la favorabilidad del mismo. Se puede determinar que no se tiene un 100% de exactitud en la implementación del algoritmo, adicionalmente varios de los tweets no fueron clasificados correctamente y esto se debe a las palabras que pueden ser usadas en la ontología, debido a que el vocabulario colombiano es tan variado y puede contener muchas más palabras que el algoritmo no tenga, se pueden ir generando más palabras en la ontología para entrenar el algoritmo.

- El modelo predictivo dio una favorabilidad al candidato actualmente electo Iván duque, estableciendo que a pesar de que han pasado varios meses una vez terminada la candidatura presidencial aún se conserva una buena imagen del mismo.
- Mediante el método implementado se identifica que la toma de tweets en tiempo real provoca que se tarde tiempo en el procesamiento de la información.
- El modelo predictivo establecido trabaja de acuerdo a los vocabularios ontológicos, realizando clasificación de mensajes con un adecuado entrenamiento.
- La arquitectura implementada fue de vitalidad debido a que el trabajo con bases de datos no relaciones y la conexión mediante herramientas como lo son Node.js generan un impacto favorable en rendimiento y desempeño del procesamiento de la información.
- La prueba de concepto permite que cualquier usuario pueda interactuar con la aplicación y obtener datos en cualquier momento.

9. TRABAJOS FUTUROS

- Implementar el modelo predictivo en otras redes sociales como Facebook o comentarios de videos en candidatos como YouTube, ya que en estas se generan comentarios políticos que pueden inducir a una posible elección.
- Implementar otro tipo de algoritmos de predicción que permitan analizar diferentes desempeños con el implementado.
- Implementar modelos predictivos para el contenido visual generado en Twitter u otras redes sociales.
- Usar la API de Twitter de versión paga, con el fin de obtener más datos y más tiempo en el uso de la misma.
- Implementar este modelo predictivo con elecciones de alcaldes u otro cargo político.
- Mejorar los predictores debido a que la muestra al tener un nivel de confianza como el mínimo, este pueda mejorar

BIBLIOGRAFIA

Cambridge Analytica, «Candidate Reputation Management PR Campaign,» 26 03 2018. [En línea]. Available: <https://ca-political.com/casestudies/casestudycolombia>. [Último acceso: 26 03 2018].

S.A.S., RESOLVE STUDIO, «MEDIMOS, ANALIZAMOS Y OPTIMIZAMOS,» 26 03 2018. [En línea]. Available: http://resolvestudio.co/elecciones_presidenciales.html. [Último acceso: 26 03 2018].

Cambridge Analytica, «Donald J. Trump for President,» 26 03 2018. [En línea]. Available: <https://ca-political.com/casestudies/casestudydonaldjtrumpforpresident2016>. [Último acceso: 26 03 2018].

El Espectador & Universidad Sergio Arboleda, «Redes sociales: ¿Cuál será su incidencia en las próximas elecciones a la Presidencia?,» El Espectador, 30 01 2018. [En línea]. Available: <https://www.elespectador.com/elecciones-2018/noticias/politica/redes-sociales-cual-sera-su-incidencia-en-las-proximas-elecciones-la-presidencia-articulo-736334>. [Último acceso: 27 03 2018].

CNN Español, «Redes sociales: donde Donald Trump tiene su verdadero poder,» 17 11 2016. [En línea]. Available: <http://cnnespanol.cnn.com/2016/11/17/redes-sociales-donde-donald-trump-tiene-su-verdadero-poder/>. [Último acceso: 27 03 2018].

CARACOL RADIO, «Les contamos las redes sociales que prefieren los colombianos,» EL HERALDO S.A., 24 08 2017. [En línea]. Available: http://caracol.com.co/programa/2017/08/25/dos_y_punto/1503616822_036506.html. [Último acceso: 27 03 2018].

R. Guns, Tracing the origins of the semantic web. Journal Of The American Society For Information Science & Technology, 2013, pp. 2173-2181.

H. B.-L. Bizer, Linked Data – The Story so far. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems, 2009.

L. F. Criado, Nosotros, los constructores de la Web Semántica. Cimientos: Open Data y Linked Open Data., Madrid, 2013.

S. Pastor, Tecnologías de la web semántica. 1ª edición. Barcelona, 1ª edición ed., UOC, 2011.

T. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Stanford Knowledge Systems Laboratory., California, 1993.

«IBM KNOWLEDGE,» [En línea]. Available:

https://www.ibm.com/support/knowledgecenter/en/SS6QYM_9.1.0/com.ibm.help.sfs.custom.doc/Fsales_SampleXMLRelatedTask.html.

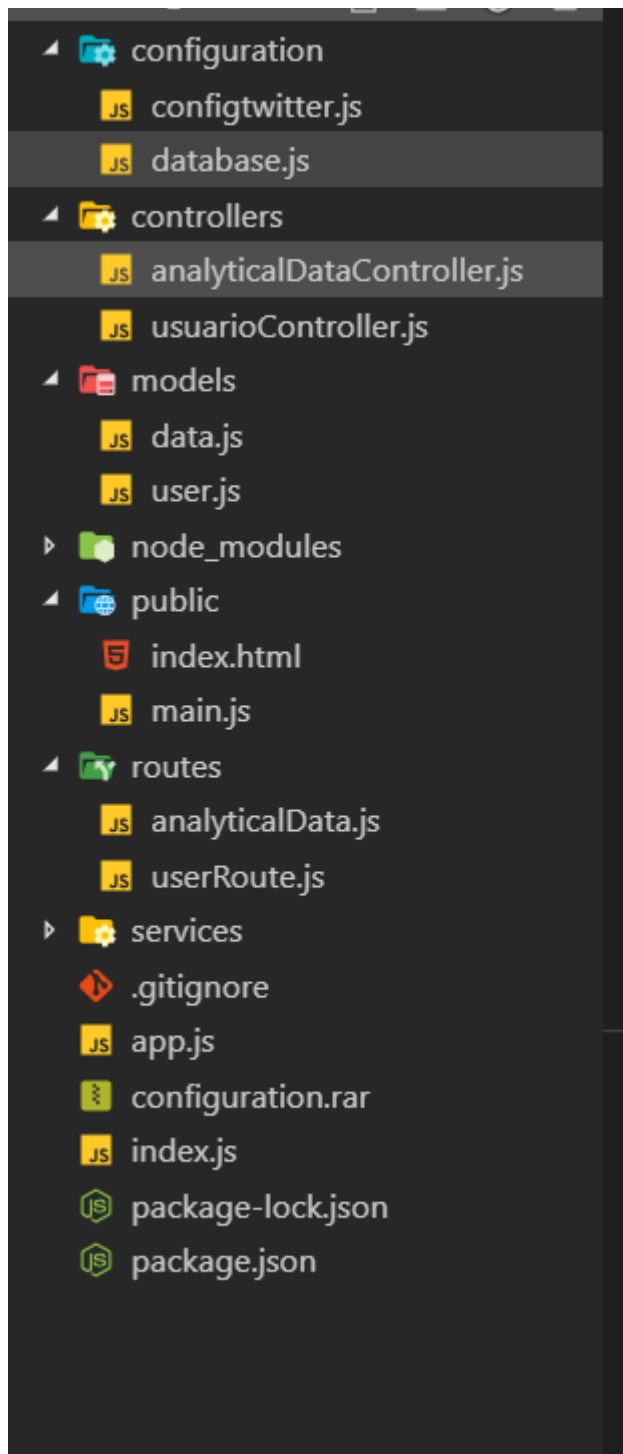
C. M. ZAPATA y G. L. y. M. J. E. GIRALDO, «UNA PROPUESTA DE METAONTOLOGÍA PARA LA EDUCACIÓN DE REQUISITOS,» [En línea].

Available: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-33052010000100004&lng=es&nrm=iso.

C. T. H. Hoang, Semantic Information Integration with Linked Data Mashups Approaches., 2014.

ANEXOS

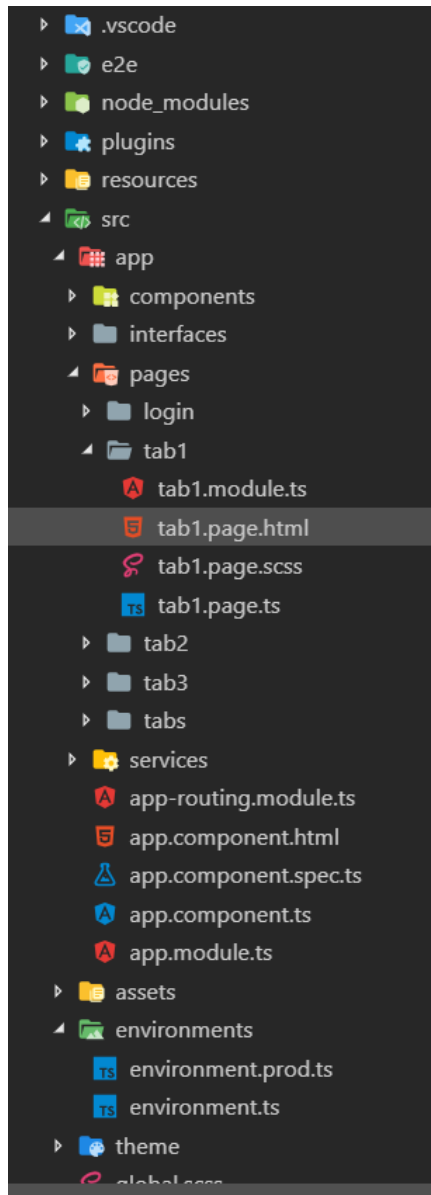
Anexo A. Componentes del BackEnd



Anexo B. Datos de entrenamiento mediante ontologías

```
1  'use strict'
2  var Twitter = require('twitter');
3  var token_twitter = require('../configuration/configtwitter');
4  var DecisionTree = require('decision-tree');
5  var server = require('../app.js');
6  var io = require('socket.io')(server);
7
8
9  var training_data = [
10   { "candidato": "@petrogustavo", "tendencia": "izquierda", "ocupacion": "Abogado", "politico": "movimiento progresista", "propuestas": "proponer una reforma constitucional", "partido": "movimiento progresista" },
11   { "candidato": "@petrogustavo", "tendencia": "izquierda", "ocupacion": "Abogado politico", "politico": "movimiento progresista", "propuestas": "proponer una reforma constitucional", "partido": "movimiento progresista" },
12   { "candidato": "@petrogustavo", "tendencia": "izquierda", "ocupacion": "Abogado politico", "politico": "movimiento progresista", "propuestas": "proponer una reforma constitucional", "partido": "movimiento progresista" },
13
14   { "candidato": "@IvanDuque", "tendencia": "derecha", "ocupacion": "Abogado politico", "politico": "centro democratico", "propuestas": "proponer una reforma constitucional", "partido": "centro democratico" },
15   { "candidato": "@IvanDuque", "tendencia": "derecha", "ocupacion": "Abogado politico", "politico": "centro democratico", "propuestas": "proponer una reforma constitucional", "partido": "centro democratico" },
16   { "candidato": "@IvanDuque", "tendencia": "derecha", "ocupacion": "Abogado politico", "politico": "centro democratico", "propuestas": "proponer una reforma constitucional", "partido": "centro democratico" },
17
18   { "candidato": "@sergio_fajardo", "tendencia": "centro", "ocupacion": "matematico politico", "politico": "alianza verde", "propuestas": "proponer una reforma constitucional", "partido": "alianza verde" },
19   { "candidato": "@sergio_fajardo", "tendencia": "centro", "ocupacion": "matematico politico", "politico": "alianza verde", "propuestas": "proponer una reforma constitucional", "partido": "alianza verde" },
20   { "candidato": "@sergio_fajardo", "tendencia": "centro", "ocupacion": "matematico politico", "politico": "alianza verde", "propuestas": "proponer una reforma constitucional", "partido": "alianza verde" },
21
22   { "candidato": "@German_Vargas", "tendencia": "derecha", "ocupacion": "economista", "politico": "partido liberal", "propuestas": "proponer una reforma constitucional", "partido": "partido liberal" },
23   { "candidato": "@German_Vargas", "tendencia": "derecha", "ocupacion": "economista", "politico": "partido liberal", "propuestas": "proponer una reforma constitucional", "partido": "partido liberal" },
24   { "candidato": "@German_Vargas", "tendencia": "derecha", "ocupacion": "economista", "politico": "partido liberal", "propuestas": "proponer una reforma constitucional", "partido": "partido liberal" },
25 ];
26
```

Anexo C. Componentes del frontend



Anexo D. Vista de ionic framework

```
6      </ion-toolbar>
7    </ion-header>
8
9
10   <ion-content>
11
12
13     <ion-card>
14
15       <ion-row ion-padding>
16         <div style="position: relative; height:200px; width:400px; margin-top:30px; margin-bottom: -47px;">
17           <canvas id="pro-chart"></canvas>
18         </div>
19       </ion-row>
20     </ion-card>
21
22     <ion-card>
23
24       <ion-row ion-padding>
25         <div style="position: relative; height:200px; width:400px; margin-top:30px; margin-bottom: -47px;">
26           <canvas id="canvas-chart"></canvas>
27         </div>
28       </ion-row>
29     </ion-card>
30
31
32     Loading...
33
34     <ion-text color="dark" text-center>
35       <h1>PERSONAJES</h1>
36     </ion-text>
37
38
39
40     <ion-row>
41       <ion-card>
42         <ion-col size="12" text-center>
43           
```

Anexo E. Manual de Usuario.



Manual_Usuario_Jere
Soft.pdf

Anexo F. Manual de instalación y configuración.



Manual_Instalacion_C
onfiguracion_Jeresoft.